REGISTRATION OF IMAGES WITH PATHOLOGIES

Xu Han

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Computer Science.

Chapel Hill
2020

Approved by:

Marc Niethammer

Martin Styner

Junier Oliva

Roland Kwitt

Gikas Mageras

ABSTRACT

Xu Han: Registration of Images with Pathologies
(Under the direction of Marc Niethammer)

Registration is one of the fundamental tasks in medical image analysis. It is an essential step for many applications to establish spatial correspondences between two images. However, image registration in the presence of pathologies is challenging due to tissue appearance changes and missing correspondences caused by the pathologies. For example, for patients with brain tumors, the tissue is often displaced by the tumors, creating more significant deformations than what is observed in a healthy brain. Moreover, a fast and accurate image registration in the presence of pathologies is especially desired for immediate assessment of the registration results.

This dissertation addresses the following problems concerning the registration of images with pathologies: (1) efficient registration between an image with pathologies and a common control atlas; (2) patient-specific longitudinal registration between pre-operative and post-recurrence images for patients with glioblastoma; (3) automatic brain extraction for images with pathologies; and (4) fast predictive registration of images with pathologies to an atlas.

Contributions presented in this dissertation are as follows: (1) I develop a joint PCA/image-reconstruction approach for images with pathologies. The model estimates quasi-normal image appearance from the image with pathologies and uses the reconstructed quasi-normal image for registration. It improves the registration accuracy compared to directly using the images with pathologies, while not requiring the segmentation of the pathological region. (2) I propose a patient-specific registration framework for the longitudinal study of tumor recurrence of patients diagnosed with glioblastoma. It models the healthy tissue appearance for each patient in the individual space, thereby improving the registration accuracy. (3) I develop a brain extraction method for images with pathologies by jointly modeling healthy brain tissue, pathologies, and non-brain volume. (4)

I design a joint registration and reconstruction deep learning model which learns an appearance mapping from the image with pathologies to atlas appearance while simultaneously predicting the transformation to atlas space. The network disentangles the spatial variation from the appearance changes caused by the pathology.

In memory of my grandparents
Qiguo and Airong

ACKNOWLEDGMENTS

First and foremost, I thank my Ph.D. advisor Dr. Marc Niethammer for his support over the past five years. It was my honor to have worked with him in image registration. Marc always has insight into the research topics that I was working on, and his patience and encouragement helped me get through many challenges and problems that I experienced. He is also a very generous and supportive mentor. Thanks to him, I was able to maintain a work-life balance during my Ph.D. journey.

Additionally, I would like to express my gratitude to my other committee members for their advice and support in my dissertation: Dr. Martin Styner, Dr. Junier Oliva, Dr. Roland Kwitt, and Dr. Gikas Mageras. As an expert in medical image analysis, Martin provided valuable feedbacks in many aspects of my research. Junier provided insights from his expertise in machine learning. Their knowledge helped strengthen the dissertation. I have enjoyed working with Roland to develop the registration framework in my first two years. He came up with many great ideas, helped me think thoroughly and design experiments to tackle the problems. Finally, I want to thank Gig for his advice on the research project that I worked on with Memorial Sloan Kettering Cancer Center. While results are not presented in this dissertation, his advice is valuable for me to understand how image registration works in clinical research and application.

I would also like to thank the following collaborators from other institutions or companies: Dr. Stephen Aylward from Kitware Inc., Dr. Spyridon Bakas from the University of Pennsylvania, and Dr. Jun Hong from Memorial Sloan Kettering Cancer Center. I enjoyed my time working with them and I appreciate their time for providing the data and valuable feedback on my research work.

I would like to thank fellow members from our UNC-Biag group: Dr. Tian Cao, Dr. Yi Hong, Dr. Xiao Yang, Dr. Heather Couture, Dr. Istvan Csapo, Zhipeng Ding, Zhenlin Xu, Zhengyang Shen, Yifeng Shi, Peirong Liu, Tian Lin, Sahin Olut. I always enjoyed our discussions and have

learned a lot from all of them. Besides my friends in my research group, I also thank all my fellow friends at UNC - Dr. Dong Nie, Danwei Li, Ming Yang, Rui Wang, and Yue Guo to name a few.

Last but not least, I thank my parents, Faming and Qin, for their love since always. I thank my grandparents, Qinguo and Airong, for helping raise me up and for the early education in my childhood. I thank my wife Sisi for her love. She sacrificed a lot of her time to take care of the family while I pursue my Ph.D. The two cute babies - Dà yǔ and Xiǎo yǔ (both of whom were born during my Ph.D.) - have brought lots of fun and happiness to our family. I also want to thank my parents in law, Jinshui and Bing, for their support and time to care for the boys.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

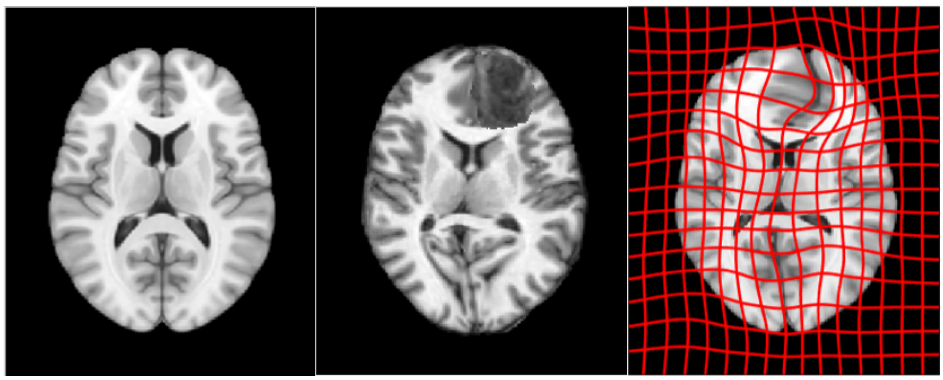
CFM	Cost Function Masking
CNN	Convolutional Neural Network
CT	Computed Tomography
DL	Deep Learning
LDDMM	Large Deformation Diffeomorphic Metric Mapping
LRS	Low-rank/Sparse
MI	Mutual Information
MRI	Magnetic Resonance Imaging
NCC	Normalized Cross Correlation
PCA	Principal Component Analysis
SSD	Sum of Squared Differences
SVD	Singular Value Decomposition
SVF	Stationary Velocity Field
TBI	Traumatic Brain Injury
TV	Total Variation

CHAPTER 1: Introduction

1.1 Motivation

Image registration is a fundamental task in medical image analysis (Brown, 1992). It aims to find a spatial transform between two images such that images after the transformation are aligned. Image registration plays an essential role in medical image applications. For example, image-to-atlas registration provides a mapping from an individual patient to an atlas. The image for each patient can then be mapped to a common coordinate system for population-based analyses (Toga and Thompson, 2001). The mapping can also be used to transfer the prior atlas information to an individual subject, for example, to obtain a segmentation (Aljabar et al., 2009). Among these applications, atlas-based brain extraction has been widely studied (Doshi et al., 2013), as brain extraction is an essential preprocessing step for many neuroimaging pipelines (Smith, 2002). Longitudinal registration (Smith et al., 2002) within patients provides a mapping for organs and tissue at different time points, which can be used to evaluate the growth and the development of the tissue of interest, e.g., tumors (Provenzale et al., 2006). It can also be used to evaluate the movements of the organs, which is required for clinical treatments (Pace et al., 2011) or image-guided radiation therapy (Lee et al., 2008).

To evaluate how images are aligned, image similarity measures are used. However, when presented with images with pathologies, registration becomes challenging as standard image similarity measures, such as the sum of squared differences (SSD) or the normalized cross-correlation (NCC), do not account for changes arising from pathologies and cannot establish reliable spatial correspondences. Directly using an image with pathologies for registration may result in unreliable deformation estimate, especially for images with large pathologies. Figure 1.1 shows such an example of registration ressult between an atlas image and an image with a brain tumor.



(a) (b) (c)

Figure 1.1: An example registration result between an atlas to an image with brain tumor: (a) atlas image; (b) image with tumor; (c) warped atlas with deformation grid. Directly using the tumor image shows crazy deformation in the tumor region.

However, registration of images with pathologies is needed, for example, to support:

1. Disease diagnosis and treatment planning using atlas-based tissue segmentation to identify traumatic brain injuries (TBI), tumors, or strokes (Irimia et al., 2012; Brock et al., 2017),

2. Treatment monitoring using longitudinal images for brain tumor recurrence assessment (Kwon et al., 2014), and

3. Automatic brain extraction from magnetic resonance brain images with pathologies (Iglesias et al., 2011), which is essential for further analysis of the brain tissue.

Additionally, as registration methods are typically formulated as optimization problems, the computational time cost can be extensive, especially for non-parametric registration methods. Thus, a fast registration method for images with pathologies is often desirable, for example, in the context of image-guided neurosurgery (Drakopoulos and Chrisochoides, 2016).

1.1.1 Efficient Registration of Images with Pathologies

A variety of approaches have been proposed to address pathological image registration. For example, cost function masking (Brett et al., 2001) and geometric metamorphosis (Niethammer et al., 2011) exclude pathological regions from measurements of image similarity. However, these approaches require prior segmentation of the pathological regions, which is a non-trivial or

labor-intensive task. It can be worse when the pathological regions are large or in anatomically critical locations. Excluding such regions and estimating the deformation field in these regions via regularization may compromise the registration accuracy (Kwon et al., 2015; Yang et al., 2016). Joint registration and segmentation approaches have also been proposed, which include estimating a latent label field to indicate missing correspondences (Chitphakdithai and Duncan, 2010; Gooya et al., 2011; Kwon et al., 2014).

A conceptually different approach is to learn normal image appearance from population data and to use it to estimate a quasi-normal image from an image with pathologies. This quasi-normal image can then be used for registration. Quasi-normal images can, for example, be estimated via a low-rank/sparse (LRS) decomposition (Liu et al., 2015) or by learning a direct mapping from an image with pathologies to a quasi-normal image via an autoencoder (Yang et al., 2016). LRS suffers from three shortcomings: First, the ideal LRS decomposition is computed based on already aligned images. Hence, in practice, registration and LRS decomposition steps need to be alternated, making the algorithm costly. Second, LRS decomposes the full population sample, causing high memory demand. Jointly with the first shortcoming, this severely limits the number of subjects that can be used for the decompositions to capture population variation. Third, while LRS reconstructs pathological areas, making them appear quasi-normal, it also blurs healthy tissue and hence may impair registration accuracy in areas unaffected by the pathology. While the autoencoder approach (Yang et al., 2016) does not blur healthy tissue and does not require alternating registration and decomposition for a full population of images, it requires a large number of training images and has so far not been extended to 3D.

Inspired by LRS and to overcome these shortcomings of LRS, I present a novel image decomposition approach that models the healthy tissue via principal component analysis (PCA) of the healthy population and the pathological region via a total variation (TV) penalty. The model registers one single image at a time thereby significantly reducing the computational cost. Chapter 3 describes the proposed method and presents the experimental results on a pseudo-tumor dataset and BraTS2015 (Menze et al., 2014; Kistler et al., 2013).

1.1.2 Patient-Specific Registration of Pre-Operative and Post-Recurrence Brain Tumor Images

Glioblastoma is one of the most common and aggressive malignant brain tumors that can significantly and heterogeneously infiltrate surrounding tissue (Omuro and DeAngelis, 2013). This infiltration complicates treatment (Price et al., 2007), as it is difficult to localize precisely the extent of infiltration. Considering that more than 80% of patients have a local tumor recurrence close to the initial resection cavity (Milano et al., 2010) (hence to the infiltrated brain tissue), accurate registration is needed between the pre-operative (pre) and the post-recurrence (post) brain tumor scans. Such registration would support research into the early detection of tumor recurrence, enabling the identification of subtle imaging phenotypic characteristics of tumor recurrence. Even though correspondences are established between longitudinal image-pairs, where scans are expected to be comparable (as they are of the same patient), registration is challenging due to there being two sources of image appearance changes: first, the pre-scans contain tumors and strong mass effect (large deformations of the normal tissue caused by the tumor); second, the post-scans typically contain tumor resection cavities (where the tumors used to be in the pre-scans) and show signs of tumor infiltration and recurrence.

While many methods, as discussed above, are aimed at dealing with images with pathologies, they do not take into account that the two scans come from the same patient. PORTR (Kwon et al., 2014) has been proposed to register between the pre-operative and the post-recurrence brain tumor scans, but it is challenging to estimate the deformations of pre scans with large tumors and post scans with significant mass effect relaxation, as PORTR excludes pathological regions, such as edema, from the matching cost. Even though its extension (Kwon et al., 2015) incorporates an inpainting approach to account for pathological regions, prior knowledge about the tumor of each scan, comprising of seed-points with associated radii and initial intensity modeling of each brain tissue type, is required for the algorithm. This manual interactive step, in addition to introducing an extra burden to the method's usability and increasing the time footprint of the method, also affects the objectivity and repeatability of the obtained results. All these together have a direct impact

on the consideration of the method for potential clinical translation, as well as for large research studies.

Therefore I present a patient-specific registration framework that models the tissue appearance in the patient space. By running PCA in the patient-specific space, the reconstructed tissue appearance better represents the given patient, which in turn improves the registration accuracy. Chapter 4 describes the patient-specific registration framework and presents the experimental results on a dataset of clinically acquired paired pre- and post-scans of patients diagnosed with de novo (primary) glioblastoma. The proposed framework does not require any manual interaction, neither in the form of segmentation nor as tumor seeding, and only requires a single modality.

1.1.3 Brain Extraction from Images with and without Pathologies

Brain extraction, commonly referred to as skull stripping, from volumetric magnetic resonance (MR) or computed tomography (CT) images (Muschelli et al., 2015) is a common preprocessing step in neuroimaging as it allows focusing the further analysis on the areas of interest. It also simplifies brain registration as a registration algorithm can focus on the relevant image parts. The most straightforward approach to brain extraction is by manual expert delineation. Unfortunately, such expert segmentations are time-consuming and very labor-intensive and, therefore, not suitable for large-scale imaging studies. Many brain extraction methods proposed are based on the registration of an atlas to the image as an initial step, where the atlas mask is used to provide coarse results. However, many challenges for automatic brain extraction remain: First, many methods show varying performances on different datasets due to differences in image acquisition (e.g., slightly different sequences or differing voxel sizes). Second, most methods only work for images that appear healthy or show minor pathologies. Strong pathologies, however, may induce large brain deformations or significant localized changes in image appearance, which can impact brain extraction. For example, for methods based on registration, the accuracy of brain extraction will depend on the accuracy of the registration, which can be severely affected when presented with pathologies.

While the decomposition model described in Chapter 3 captures the pathological region via a TV term, it does not model the non-brain regions, such as the skulls. The skull is, for example,

usually a thin, shell-shape structure, and other non-brain tissue may be irregularly shaped with various intensities. I extend the decomposition framework to capture the non-brain tissue as well. The extended framework decomposes an image into three parts: healthy brain tissue, pathologies, and non-brain region. Chapter 5 describes the proposed brain extraction framework in detail and presents experiments on two datasets of normal images, one brain tumor dataset, and one clinically acquired TBI dataset.

1.1.4 A Deep Network for Joint Registration and Reconstruction of Images with Pathologies

Traditionally, image registration is formulated as an optimization problem seeking to minimize the dissimilarity between a warped source image and a target image while simultaneously encouraging spatially regular transformations. To capture large deformations, fluid-based registration models are frequently used (Modersitzki, 2004), e.g., stationary velocity field (SVF) (Vercauteren et al., 2009) or large deformation diffeomorphic metric mapping (LDDMM) approaches (Beg et al., 2005), which can guarantee diffeomorphic transformations if sufficiently regularized. Non-parametric image registration models (Modersitzki, 2004), such as SVF and LDDMM, require optimizing over millions of parameters in 3D, which is usually very slow. Hence, deep learning (DL) approaches have been proposed for such registration models (Yang et al., 2017; Balakrishnan et al., 2019; Shen et al., 2019a,b). By shifting the computational cost to the training time, DL approaches are orders of magnitudes faster at test time than numerical optimization, while retaining registration accuracy.

While many registration approaches for normal images or images with similar appearance have been proposed, a limited body of literature exists for the registration of images with pathologies, especially DL approaches. Existing approaches for registration of images with pathologies, as well as their limitations, have been discussed above. Reconstruction of quasi-normal appearance is attractive for registration of images with pathologies, as it does not require a prior segmentation and can establish correspondence in the pathological region. Tumor-to-quasi-normal appearance can be learned via quasi-lesions with a variational autoencoder (Yang et al., 2016), or from a statistical model of a healthy population (Liu et al., 2014). The quasi-lesion approach (Yang et al., 2016)

introduces synthetic tumors and learns to reconstruct the underlying healthy appearance. However, the resulting reconstructions are still subject to mass effects and therefore do not correctly separate appearance learning from such deformation changes. Existing approaches based on statistical models (Liu et al., 2014, 2015), including the model proposed in Chapter 3, require underlying registrations to a common space for quasi-normal image reconstruction. Nevertheless, as a good alignment in cases of mass effect cannot be obtained without a good reconstruction, registration and reconstruction need to be interleaved in a costly iterative scheme.

Inspired by a work on shape and appearance disentangling (Shu et al., 2018), I propose a deep neural network to simultaneously register a brain tumor image to an atlas while reconstructing a quasi-normal image *in atlas space*. The reconstructed quasi-normal image is, in turn, used in the similarity loss to guide the network to learn the spatial transformation from the image to the atlas. Chapter 6 presents the proposed deep learning model, as well as the experiments on a pseudo-tumor dataset and a real brain tumor dataset.

1.2 Thesis Statement and Contributions

Thesis: *Advanced mathematical models can efficiently and effectively estimate normal image appearance from images with pathologies, thereby helping to improve registration accuracy. Furthermore, a deep regression model allows for fast and accurate registrations of such images.*

The following contributions are presented for this thesis:

1. I propose an image decomposition model that reconstructs a quasi-normal image from an image with pathology. The quasi-normal image is then used for registration.

2. I develop a patient-specific strategy that models the healthy tissue of a patient in the patient space. This patient-specific strategy allows for more accurate modeling of the quasi-normal reconstruction for each patient, thereby resulting in improved longitudinal registration accuracy.

3. I propose a brain extraction framework that is designed for images with pathologies by jointly modeling the healthy brain issue, pathologies, and non-brain volume. While it is designed for

brain image that contains pathologies, it shows improved brain extraction results in datasets with and without pathologies.

4. I develop a deep learning model that jointly learns the registration of an image with pathologies to an atlas and appearance changes between the image with pathologies and a quasi-normal image. As the network disentangles the spatial variation from appearance learning, it results in more accurate registration results.

1.3 Overview of Chapters

This thesis is organized as follows: Chapter 2 introduces background for image registration, principal component analysis (PCA), and deep learning (DL). Chapter 3 introduces the decomposition model for an image with pathologies. Chapter 4 presents the patient-specific strategy for the longitudinal registration between pre-operative and post-recurrence brain tumor scans. Chapter 5 extends the decomposition model and presents a brain extraction framework for brain images with pathologies. Chapter 6 proposes a deep learning method for joint registration and reconstruction of an image with pathologies. Chapter 7 concludes this thesis with a discussion of the contributions and an outlook on future work.

CHAPTER 2: Background

This chapter presents some essential background knowledge that is relevant to this dissertation. Specifically, I will first introduce the topic of image registration in Chapter 2.1 as well as three existing basic methods that aim to register images with pathologies. Principal component analysis (PCA) is another essential topic for this dissertation, which will be presented in Chapter 2.2. It includes some mathematical background and how the eigendecomposition and the singular value decomposition (SVD) can be used to perform the PCA. Finally, chapter 2.3 presents some basics for deep learning, which includes convolutional neural networks and autoencoders, as well as how they connect to the PCA. Three popular deep learning-based image registration models are also briefly summarized in chapter 2.3.

2.1 Image Registration

Image registration is a process to find spatial correspondences between images. These images can be taken at different time-points, by different machines, or in different modalities (Zitova and Flusser, 2003). Aligning these images into a common coordinate system provides convenience and is necessary for comparison and analysis between these images, as corresponding structures in the images should have similar appearances and functions. While image registration can be performed between two or more images, in this dissertation, I will focus on registration between two images.

In image registration, one image is referred to as the source image or the moving image, and the other one is the target image or the fixed image. Given a moving image I_M and a target image I_T, an image registration aims to find a spatial transformation Φ that deforms the moving image, such that the deformed moving image is similar to the target image, based on some similarity criterion. Additionally, it would be desirable that the transformation map has some nice properties, for example, to be smooth or diffeomorphic (Ashburner, 2007). Many methods have been proposed

for image registration (Modersitzki, 2004), and typically these methods formulate the problem as an optimization problem, which aims to minimize the following energy function:

$$E(\mathbf{\Phi}) = Reg[\mathbf{\Phi}] + \frac{1}{\sigma^2} Sim[I_M \circ \mathbf{\Phi^{-1}}, I_T], \tag{2.1}$$

where σ is a weight balancing two terms. The term $Reg[\cdot]$ penalizes the irregularity of the transformation, and the term $Sim[\cdot]$ penalizes the dissimilarity between the warped moving image and the target image.

2.1.1 Image Similarity Measure

The image similarity term $Sim[\cdot]$ is used to evaluate how well the images are aligned after the registration, i.e., how well the correspondence between the images is established. Typically the correspondence should be matched between manually annotated landmarks or regions such as the segmentation of regions of interest. However, obtaining such landmarks or segmentation is labor-intensive and time-consuming. Therefore, image intensities are commonly used to evaluate similarity measures, and standard image similarity measures include the sum of squared differences (SSD), the normalized cross-correlation (NCC), and mutual information (MI).

Sum of squared differences. The simplest way to compare two images I_1 and I_2 is to compute the SSD between the intensities of the image pair

$$SSD(I_1, I_2) = \int_\Omega (I_1(x) - I_2(x))^2 \, dx, \tag{2.2}$$

where Ω is the image spatial domain. SSD is a popular measure when two images are of the same modality and have a similar intensity range.

Normalized cross correlation. NCC is defined as the ratio between the cross-covariance of two images and the multiplication of the standard deviation of each image:

$$NCC(I_1, I_2) = \frac{\int_\Omega (I_1(x) - \bar{I_1})(I_2(x) - \bar{I_2}) \, dx}{\sqrt{\int_\Omega (I_1(x) - \bar{I_1})^2 \, dx} \sqrt{\int_\Omega (I_2(x) - \bar{I_2})^2 \, dx}} \in [-1, 1], \tag{2.3}$$

where $\bar{I}$ refers to the mean intensity of the image I. NCC between two images is ± 1 when two images are statistically dependent, and is 0 when independent. When using NCC as the similarity measure for registration, people aim to maximize the statistical dependence of the aligned image, and therefore $1 - NCC(I_1, I_2)^2$ is typically used in the energy function. Moreover, any linear relationship between intensities will lead to a perfect match (1 or -1).

Mutual Information. MI is a more flexible similarity measure than NCC, coming from information theory. Image intensities are viewed as random variables. It is to evaluate, given the intensity of one image, how well one can predict the intensity of the other image. Given two random variables X and Y, the mutual information is defined as

$$MI(X, Y) = \int \log \frac{p(x, y)}{p(x)p(y)} \, dxdy, \tag{2.4}$$

where $p(x)$ and $p(y)$ are probability density function of X and Y, and $p(x, y)$ is the joint probability density function. To compare two images, $p(x, y)$ is the normalized joint intensity histogram. When two images are independent, $p(x, y) = p(x)p(y)$ and therefore the MI is zero. The more the image pairs are similar, the higher the value of MI is.

2.1.2 Regularization Energy

The regularization term $Reg[\cdot]$ penalizes the spatial irregularity of the transformation, which encourages a plausible transformation. How to choose a proper regularization depends on the application of the image registration, which leads to different deformation models. For example, to register between rigid objects, such as bones, a rigid transformation should be enough, while an elastic model or a fluid model may be more suitable to evaluate local deformation of the soft tissue. Depending on whether the deformation field can be parametrized, image registration can be categorized into parametric and non-parametric models.

2.1.2.1 Parametric models.

Parametric registration models include many low-dimensional transformations, such as similarity, rigid, and affine transformations. The regularization term usually is omitted from the

equation 2.1, as it is constrained via the low degrees of freedom of the chosen transformation model. Among them, affine registration is frequently used as an initialization step for image registration, as most of the complex registration models are affected by affine initialization[1]. B-spline (Rueckert et al., 1999) registration is another type of parametric model, which parametrizes the deformation field via a grid of control points. The deformation field is obtained by b-spline interpolation of the control points.

2.1.2.2 Non-parametric models.

In non-parametric models, registration is parametrized by functions, e.g. via displacement field, velocity field or vector momentum fields. In such models, the deformation field is given by $u(x) = \Phi(x) - x$, where $\Phi(x)$ is the transformation map, and regularization of the deformation field is necessary for non-parametric models to avoid implausible deformations or to ensure the physical properties of the transformation. The most commonly used non-parametric models are the physically motivated elastic and fluid registrations, but other models such as diffusion and curvature are also widely used.

Elastic Registration. In elastic registration (BROIT, 1981), images are viewed as deformable elastic bodies, and the displacement u is derived based on a linear elasticity model. The regularization energy for elastic registration is

$$Reg[u] = \int_{\Omega} \frac{\mu}{4} \sum_{j,k=1}^{d} (\partial_{x_j} u_k + \partial_{x_k} u_j)^2 + \frac{\lambda}{2} (\text{div } u)^2 \, dx, \tag{2.5}$$

where μ and λ are the Lamé constants that control the elastic behavior, d is the image dimension, and Ω is the image domain. While the elastic model is appropriate for registration arising from an elastic material, it may not be sufficient to capture deformation between dissimilar objects as it implicitly assumes a small displacement field. In such a case, fluid registration may be a proper choice.

[1]Curvature registration is one exception, which is invariant to affine transformations.

Fluid Registration. The key difference between fluid registration and elastic registration is the regularization energy term (Modersitzki, 2004). While the regularization is based on the displacement field u for elastic models, in fluid-based registration models (Wollny and Kruggel, 2002), it is based on a velocity field $v = \partial_t u$, which is the derivative of the displacement field with respect to time t. Registration for fluid-based models can then be written as

$$\begin{aligned} E(v) &= \int_0^1 \|v\|_L^2 \, dt + \frac{1}{\sigma^2} Sim[I_M \circ \Phi^{-1}(1), I_T], \\ \text{s.t.} \quad & \Phi_t^{-1} + D\Phi^{-1} v = 0, \quad \text{and} \quad \Phi^{-1}(0) = id. \end{aligned} \tag{2.6}$$

Here D denotes the Jacobian, $\|v\|_L^2 = \langle L^\dagger L v, v \rangle$, where L is a self-adjoint differential operator and $L^\dagger$ is the adjoint. The first constraint is the advection equation that transports the deformation map based on the velocity field v and the second constraint indicates that at time $t = 0$, the transformation is the identity map. In large deformation diffeomorphic metric mapping (LDDMM) (Beg et al., 2005), the sought-for velocity field is time and spatial dependent, while a simpler model is the stationary velocity field (SVF), where the velocity field is temporal-invariant (Modat et al., 2012). Diffeomorphism is theoretically guaranteed for fluid-based registration when the velocity field is sufficiently smooth (Dupuis et al., 1998). Additionally, fluid-based registration can be parametrized by a vector momentum field $m = L^\dagger L v$. The vector momentum-parametrized SVF (vSVF) (Shen et al., 2019a; Niethammer et al., 2019), for example, can be formulated as the following:

$$\begin{aligned} E(m) &= \langle m_0, v_0 \rangle + \frac{1}{\sigma^2} Sim[I_M \circ \Phi^{-1}(1), I_T], \\ \text{s.t.} \quad & \Phi_t^{-1} + D\Phi^{-1} v = 0, \quad \Phi^{-1}(0) = id, \quad \text{and} \quad v_0 = (L^\dagger L)^{-1} m_0 \end{aligned} \tag{2.7}$$

Diffusion Registration. While both elastic and fluid registrations are physically motivated, the diffusion model is motivated to smooth the displacement field (Modersitzki, 2004). The regularizer of the diffusion model is simply the gradient of the displacement, component by component (Fischer and Modersitzki, 1999):

$$Reg[u] = \frac{1}{2} \sum_{i=1}^{d} \int_\Omega \|\nabla u_i\|^2 \, dx. \tag{2.8}$$

One particular advantage of the diffusion model is the spatial decoupling of the regularization, which allows for a fast numerical solution and makes it attractive (Modersitzki, 2004).

Curvature Registration. Elastic, fluid, and diffusion registrations are variant to affine transforms, which means that an affine initialization is required for these models and that they are affected by different affine initialization results. Hence the curvature registration is proposed to circumvent this problem (Fischer and Modersitzki, 2003). Particularly, the regularization of the curvature registration is on the second-order derivative of the displacements, i.e. the Laplacian of the displacements:

$$Reg[u] = \frac{1}{2} \sum_{i=1}^{d} \int_{\Omega} (\Delta u_i)^2 \, dx. \tag{2.9}$$

Because of the Laplacian operator on the displacement field, the registration results are invariant to affine transformation and the transformation is smoother than diffusion registration, which makes the curvature registration an attractive model (Modersitzki, 2004).

2.1.3 Image registration in the presence of pathologies

Registration of images with pathologies is challenging due to missing correspondence between the healthy image and the image with pathology. In section 1.1.1, I mentioned several existing methods that aim to register images with pathologies. Based on how they address the missing correspondence, these methods can be classified into three categories: methods that ignore the pathologies, methods that model the pathologies, and methods that remove the pathologies. I will briefly discuss three representative methods: cost function masking (Brett et al., 2001), tumor growth modeling (Gooya et al., 2012) and the low-rank/sparse (LRS) approach (Liu et al., 2014)

Cost Function Masking. Cost function masking (CFM) (Brett et al., 2001) was proposed in studies of patients with a focal brain lesion and has been a very popular method when registering an image with pathology. As image similarity measures, such as SSD or NCC, cannot handle the missing correspondence between the pathological region and the healthy region, CFM restricts the image similarity measure outside the pathological region. Then instead of being calculated over the image spatial domain in equation (2.2) and (2.3), SSD and NCC are calculated only over the healthy

region. The deformation inside the pathological region is then interpolated by surrounding results to minimize the regularization penalty and, therefore, is not affected by the appearance of the brain lesion.

Tumor Growth Model. Tumor growth model is used in a joint registration and segmentation method (Gooya et al., 2010, 2011, 2012) to study patients with a Glioblastoma (GBM). GBM is one of the most common and malignant brain tumors. It varies in shape and size, and it infiltrates the surrounding healthy tissue and causes deformation, aka mass effect. As a result, registering an image with a GBM is challenging, as it usually has large deformations, in addition to the missing correspondence casued by the tumor. When registering a tumor image to an atlas, the tumor growth model induces a tumor inside the atlas and simulates the biophysical growth of the tumor via a diffusion-reaction equation (Hogea et al., 2008). Such simulated growth of a tumor in a healthy image reduces the structural differences between two images and hence establishes correspondence between images.

Low-rank/Sparse. Another way to register images with pathologies is by learning a mapping from abnormal appearance to quasi-normal appearance. The quasi-normal image appearance reconstruction was first obtained via a low-rank/sparse decomposition (Wright et al., 2009) and was integrated into an iterative registration framework (Liu et al., 2014, 2015) that aims to register a group of images with large deformation and pathologies. Assuming that pathologies generally do not occur at the similar locations and do not have a similar appearance, the low-rank component of the population represents information that is consistent across patients, i.e., quasi-normal appearance, and the sparse component captures the rest (information that is not consistent), i.e., abnormal appearance. I use the term "quasi-normal" to describe the reconstructed image because it is pathology-reduced or pathology-free, and as it is learned from a population, it is not a real normal image of the patient. The reconstruction of the quasi-normal appearance is more accurate if images are registered to a common coordinate system, e.g., an atlas, and the registration to the atlas benefits from a reasonable reconstruction of the quasi-normal image. Therefore, the decomposition and

registration are formed in an iterative scheme, and the framework outputs a quasi-normal image as well as a deformation of the image to the atlas at the convergence.

2.2 Principal Component Analysis

2.2.1 Introduction

Principal Component Analysis (PCA) is a technique to reduce the dimensionality of a data-set consisting of a set of correlated variables while retaining as much as possible of the variations in the data-set (Jolliffe, 2003). It is achieved by a vector space transform - an orthogonal transformation - to reduce the dimensionality of the data-set. It projects the original data-set to a lower-dimension space and interprets it in a few variables. These variables are called *Principal Components*. PCA is a very common dimensionality reduction technique for many applications in computer vision (De la Torre and Black, 2001), such as face recognition (Draper et al., 2003) and image compression (Clausen and Wechsler, 2000). PCA is also popular in medical imaging community for applications such as image fusion (He et al., 2010), image segmentation (Li et al., 2006), and image registration (Huizinga et al., 2016). Two popular algorithms to perform the PCA are through the eigendecomposition of the sample covariance matrix and the singular value decomposition of the data matrix (Richardson, 2009).

2.2.2 Eigendecomposition

In this section, I will show how the PCA can be achieved by an eigendecomposition of the covariance matrix. Assume I have a data-set has m variables and n observations. For image application, this usually means n images, each with m pixels/voxels. The data-set can be represented as a matrix $\boldsymbol{X}$ with a size of $m \times n$. Also I assume the matrix is centered, i.e., the column means have been subtracted from the matrix and are now zero. I am looking for an linear transform on $\boldsymbol{X}$, such that the variables of the new data-set $\boldsymbol{Y}$ are uncorrelated. The covariance matrix is used to measure how well two variables are correlated.

Assuming the transformation is $\boldsymbol{T}$ and $\boldsymbol{Y} = \boldsymbol{TX}$, the covariance matrix $\boldsymbol{C}$ for $\boldsymbol{Y}$ is first computed, which is given by

$$\boldsymbol{C} = \frac{1}{n-1}\boldsymbol{YY}^T = \frac{1}{n-1}\boldsymbol{TXX}^T\boldsymbol{T}^T = \frac{1}{n-1}\boldsymbol{T}(\boldsymbol{XX}^T)\boldsymbol{T}^T. \tag{2.10}$$

Matrix $\boldsymbol{S} = \boldsymbol{XX}^T$ is symmetric and every square symmetric matrix $\boldsymbol{S}$ can be orthogonally diagonalized as

$$\boldsymbol{S} = \boldsymbol{EDE}^T, \tag{2.11}$$

where $\boldsymbol{E}$ is an $m \times m$ orthonormal matrix whose columns are the eigenvectors of $\boldsymbol{S}$, and $\boldsymbol{D}$ is a diagonal matrix whose diagonal entries are the eigenvalues of $\boldsymbol{S}$. Moreover, as PCA is trying to convert the data-set to a set of uncorrelated variables, which means the covariance matrix of the new data-set should be a diagonalized matrix. Thus the covariance matrix can be rewritten as

$$\boldsymbol{C} = \frac{1}{n-1}\boldsymbol{TST}^T = \frac{1}{n-1}\boldsymbol{T}(\boldsymbol{EDE}^T)\boldsymbol{T}^T. \tag{2.12}$$

Therefore if $\boldsymbol{T} = \boldsymbol{E}^T$, the new covariance matrix becomes

$$\boldsymbol{C} = \frac{1}{n-1}\boldsymbol{D}. \tag{2.13}$$

The rows of $\boldsymbol{T}$ are called principal components (eigenvectors of $\boldsymbol{S}$). Additionally, the principal components can be ordered based on its corresponding eigenvalues, and the largest variance corresponds to the first principal component and the second largest variance corresponds to the second principal component, etc.

2.2.3 Singular Value Decomposition

Here I show how the singular value decomposition (SVD) can be used for PCA (Richardson, 2009). The SVD is a very important tool for PCA in medical imaging application, although the eigendecomposition can also be used theoretically. First of all, as the number of voxels m can be

very large, for example one magnetic resonance image can have a size of $m = 100 \times 100 \times 100$, which results in 1 million parameters, the covariance matrix will be huge as it has a size of $m \times m$. Hence the eigen-decomposition of the covariance matrix is practically impossible. On the other hand, the number of observations is usually a small number. It makes sense to perform the SVD on a much smaller matrix $\boldsymbol{X} \in \mathcal{R}^{m \times n}$. Secondly, the SVD is more preferred for rank-deficient and ill-conditioned matrices in general and has better numerical stability than the eigendecomposition (DO Q, 2012). Given a matrix $\boldsymbol{A} \in \mathcal{R}^{n \times m}$, the SVD of $\boldsymbol{A}$ results in

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T, \tag{2.14}$$

where $\boldsymbol{U}$ and $\boldsymbol{V}$ are orthonormal matrices with a size of $n \times n$ and $m \times m$ respectively, $\boldsymbol{\Sigma} \in \mathcal{R}^{n \times m}$ is a diagonal matrix and its entries σ_i are non-negative values, which are called singular values of $\boldsymbol{A}$. Now I will connect the SVD to PCA. Given a matrix $\boldsymbol{A} \in \mathcal{R}^{n \times m}$, with its SVD, I have

$$\begin{aligned} \boldsymbol{A}^T\boldsymbol{A} &= (\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T)^T(\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T) \\ &= \boldsymbol{V}\boldsymbol{\Sigma}^{\boldsymbol{T}}\boldsymbol{U}^T\boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^T \\ &= \boldsymbol{V}(\boldsymbol{\Sigma}^T\boldsymbol{\Sigma})\boldsymbol{V}^T. \end{aligned} \tag{2.15}$$

Since $\boldsymbol{V}$ is an orthonormal matrix, the eigenvalues of $\boldsymbol{A}^T\boldsymbol{A}$ are the squares of the singular values of $\boldsymbol{A}$. In addition, the eigenvectors of $\boldsymbol{A}$ are columns of $\boldsymbol{V}$, which will be used to derive how we use the SVD for the PCA.

Assuming again the data-set is written as a matrix $\boldsymbol{X}$ (centered, i.e., column means have been subtracted) with a size of $m \times n$, a new matrix $\boldsymbol{Z} \in \mathcal{R}^{n \times m}$ can be defined , and

$$\boldsymbol{Z} = \frac{1}{\sqrt{n-1}}\boldsymbol{X}^T. \tag{2.16}$$

Then, I have

$$\begin{aligned} \boldsymbol{Z}^T\boldsymbol{Z} &= (\frac{1}{\sqrt{n-1}}\boldsymbol{X}^T)^T(\frac{1}{\sqrt{n-1}}\boldsymbol{X}^T) \\ &= \frac{1}{n-1}\boldsymbol{X}\boldsymbol{X}^T. \end{aligned} \tag{2.17}$$

Notice that $\boldsymbol{Z}^T\boldsymbol{Z}$ is the covariance matrix of $\boldsymbol{X}$. Therefore if I perform the SVD of the matrix $\boldsymbol{Z}$, the principal components will be the columns of the orthonormal matrix $\boldsymbol{V}$. Finally I can project the original data $\boldsymbol{X}$ with the directions defined by the principal components $\boldsymbol{V}$ and obtain a new set of data

$$\boldsymbol{Y} = \boldsymbol{V}^T\boldsymbol{X}. \tag{2.18}$$

2.2.4 Dimensionality Reduction

I discussed the technical detail of PCA and how the eigendecomposition and the SVD can be used for PCA in previous sections. Here I will briefly discuss how the PCA can be used for dimensionality reduction. From the previous section, after I perform the SVD of $\boldsymbol{Z} = \frac{1}{\sqrt{n-1}}X^T$, $\boldsymbol{V}$ is of size $m \times m$ and the columns of $\boldsymbol{V}$ are principal component directions. The SVD sorts the principal components based on their singular values (i.e., variance of the covariance matrix) in decreasing order. Therefore the matrix $\boldsymbol{V}$ can be truncated and only the first $r < m$ columns can be kept, where r is a chosen number. Typically, an appropriate r is selected such that a minimal cumulative percentage of the variance is retained (Wilks, 2011), e.g., 70% or 80%. Let the truncated matrix be $\tilde{\boldsymbol{V}} \in \mathcal{R}^{m\times r}$. The new data is $\tilde{\boldsymbol{Y}} = \tilde{\boldsymbol{V}}^T\boldsymbol{X} \in \mathcal{R}^{r\times n}$. It has a dimension of r which can be $\ll m$. In addition, variables in Y are uncorrelated by construction. The dimension of the original data can be recovered by $\tilde{\boldsymbol{X}} = \tilde{\boldsymbol{V}}\tilde{\boldsymbol{Y}}$. Here although $\tilde{\boldsymbol{X}}$ and $\boldsymbol{X}$ have the same dimension, they are not the same because the matrix $\boldsymbol{V}$ has been truncated, i.e., $\tilde{\boldsymbol{V}}\tilde{\boldsymbol{V}}^T$ is not identity matrix. However, $\tilde{\boldsymbol{X}}$ keeps the most important information of $\tilde{\boldsymbol{X}}$ as it goes through the dimensionality reduction procedure.

Moreover, among all the transformations $\boldsymbol{T}$ with r columns, $\boldsymbol{V}^T$ maximizes the variance in the original data that has been preserved, while minimizing the reconstruction error of $\|\boldsymbol{X} - \tilde{\boldsymbol{X}}\|_F^2$, where $\|\cdot\|_F$ denotes the frobenius norm (Turk and Pentland, 1991).

2.3 Deep Learning

Deep learning is a class of machine learning methods that were inspired by the structure and the function of the human brain (Goodfellow et al., 2016). It has become very popular in machine learning, computer vision (Voulodimos et al., 2018), natural language processing (Young et al., 2018), and medical image application (Litjens et al., 2017). The word "deep" here refers to the number of layers in the network. They learn from the training data to extract abstract information of the input. Fig 2.1 shows a very simple network structure with three hidden layers. Layers shown in the figure are called fully connected layers, as each node (all called neuron) in the next layer is connected to all the neurons in the previous layers.

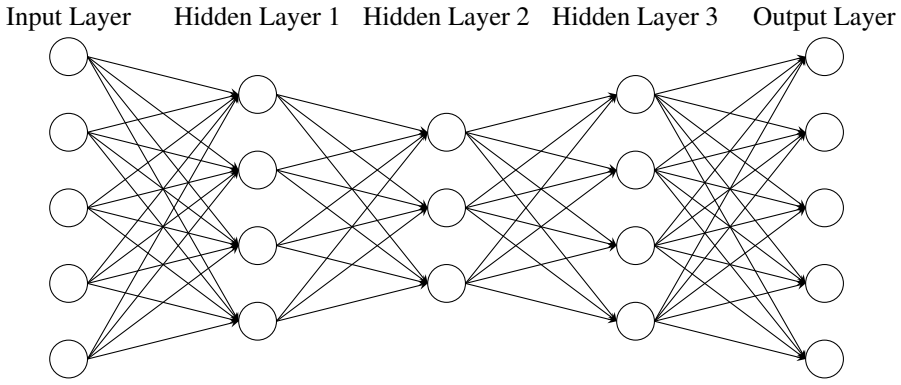


Figure 2.1: A deep learning diagram including an input, an output and three hidden layers. In practice, there are usually many hidden layers.

Mathematically, we can write a single layer of the network as follows:

$$y = \sigma(W^T x + b). \tag{2.19}$$

where W and b are the weights for the given layer, $\sigma(\cdot)$ is a non-linear function, commonly referred to as the activation function. With multiple hidden layers stacked together, we have the following

equation for a deep network with n layers:

$$y = \sigma_n(W_n^T \sigma_{n-1}(W_{n-1}^T \sigma_{n-2}(...\sigma_1(W_1^T x + b_1)...) + b_{n-1}) + b_n). \tag{2.20}$$

During training, the network optimizes over the weights with respect to a loss function, which typically is how the given problem is modeled. These weights are then fixed during the test stage for inference. How to choose a proper loss function is highly related to the problems. For example, for a regression problem where a real-value is predicted, the mean squared error (MSE) is usually used, while for a classification problem, the cross-entropy (CE) loss is commonly used. Activation functions are essential for a network to learn a non-linear mapping between the input and the output. Without activation functions, the equation above is nothing else than a combination of several linear functions, which results in a linear mapping. Several most popular types of activation functions are Sigmoid, Tanh, and ReLu, as well as their variants (Nwankpa et al., 2018).

2.3.1 Convolutional Neural Networks

Convolutional layers (LeCun et al., 1989) are one of the most popular layers in deep learning, especially in imaging applications. This is due to the high dimensional input, which makes it impractical to solely use fully connected layers. The weights for a convolutional layer are a set of learnable filters (Goodfellow et al., 2016). Each filter is separately applied to the input volume (using the dot product) to obtain a response. This can be viewed as a convolutional operation on the input layer, while the filters are the convolutional kernels. All responses are concatenated together to generate the output volume. The advantage of using convolutional layers is parameter sharing and local connectivity, which keeps the spatial dependencies of the data while dramatically reducing the number of parameters in one layer. In addition to the filter size, the stride and the padding also need to be specified. Paddings are often used when a particular output volume size is desired. Typically zero padding is used which adds an appropriate number zero values to the boundary of the input layer before it is convolved. Stride controls how the filter convolves with the input volume. The filter can be shifted one step at a time (stride 1) or more than one step. This will also result in

different output sizes. When the stride is greater than 1, the convolutional layer has another usage, i.e., pooling (downsampling). Pooling layers are used to reduce the spatial size of the representation and, thereby, the parameters in a convolutional neural network. The most commonly used pooling layer is the *max-pooling*. Others being used are the *average-pooling* and the $\mathcal{L}_2$*-pooling*.

2.3.2 Autoencoder

One example of a deep learning network that connects back to the previous PCA section is the Autoencoder. The goal of an autoencoder is similar to the PCA but is using deep networks to learn a representation of the data-set (Goodfellow et al., 2016), typically in a much lower dimension. The network takes the original data, $\boldsymbol{X}$, as the input and tries to reconstruct the input data as its output, $\boldsymbol{Y}$. The network is trained to minimize the reconstruction loss between the output and input, $\mathcal{L}(\boldsymbol{X}, \boldsymbol{Y})$. Usually, the path from the input layer to the representation layer is called an encoder, which can be represented as a function, $\boldsymbol{Z} = f(\boldsymbol{X})$, where $\boldsymbol{Z}$ is the latent representation. The path from the representation layer to the output layer is called a decoder, which can be written as a function, $\boldsymbol{Y} = g(\boldsymbol{X})$. Thus we have $\boldsymbol{Y} = g(f(\boldsymbol{X}))$. If the representation layer has a very large capacity, the network will simply copy the input to output, i.e., learn an identity map, which is not what it is aimed for. Therefore, a key point here is to make the representation layer a "bottleneck", which forces the network to learn a lower-dimensional representation of the data.

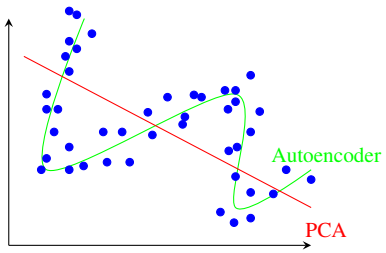


Figure 2.2: Difference between PCA and autoencoder. Blue points are example data points. Red line is what PCA can learn. Green curve is what an autoencoder can learn.

If activation functions are not used, and the MSE is chosen as the reconstruction loss function, then the autoencoder is equivalent to the PCA. Because of the activation functions, a neural network can learn a non-linear relationship. Therefore, an autoencoder can be viewed as a generalization of the PCA (non-linear PCA). Fig. 2.2 shows an example of the difference between the PCA and an

autoencoder. Assuming the data points are shown as blue points. The PCA can only learn a linear relationship in the data, thereby may not properly represent the data. However, with non-linear activation functions, an autoencoder can learn a good representation of the data in this particular case.

The standard autoencoder mentioned above is usually called an under-complete autoencoder because of the low-dimension of the latent space (Goodfellow et al., 2016). Problems occur if the latent space has a higher dimension, as mentioned before. In such a case, regularizations are added to the autoencoder to avoid learning an identity map, which leads to many different designs of the regularized autoencoder. Several examples are as follows:

- *Sparse Autoencoder.* To avoid learning an identity map, a sparse autoencoder (SAE) (Goodfellow et al., 2016) adds a sparsity penalty on the latent space $\Omega(\boldsymbol{Z})$, in addition to the reconstruction errors. An SAE is commonly used for a classification problem, as the regularization forces the network to respond to the unique statistical features of the training dataset.

- *Denoising Autoencoder.* Instead of learning to reconstruct the original input data, a denoising autoencoder (DAE) (Goodfellow et al., 2016) adds "noise" to the input data and feeds the data with noise $\tilde{\boldsymbol{X}}$ as the input and then learn to reconstruct the original data. As the input to the DAE is the corrupted data, it must revert this corruption instead of simply copying the input to the output.

2.3.3 Variational Autoencoder

In machine learning, a generative model is an unsupervised technique to learn a data distribution. It then can be used to generate new data points by sampling from the learned distribution. Two most commonly used generative models are variational autoencoders (VAE) (Kingma and Welling, 2013) and generative adversarial networks (GAN) (Goodfellow et al., 2014). In this section, I will present VAE, which is a special type of autoencoder.

Assume the data variable $\boldsymbol{x}$ is generated from a latent variable $\boldsymbol{z}$, which has a prior distribution $p(\boldsymbol{z})$. The posterior distribution of $p(\boldsymbol{z}|\boldsymbol{x})$ is given by Bayes' rule, i.e.,

$$p(\boldsymbol{z}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}, \boldsymbol{z})}{p(\boldsymbol{x})} = \frac{p(\boldsymbol{x}|\boldsymbol{z})p(\boldsymbol{z})}{\int p(\boldsymbol{x}|\boldsymbol{z})p(\boldsymbol{z})\, d\boldsymbol{z}}. \tag{2.21}$$

However, the integral of the marginal likelihood $p(\boldsymbol{x})$ is typically intractable, making the posterior distribution also intractable. A VAE approximates it via a tractable distribution $q_\phi(\boldsymbol{z}|\boldsymbol{x})$. In VAE, the encoder and decoder can be viewed as conditional distributions $q_\phi(\boldsymbol{z}|\boldsymbol{x})$ and $p_\theta(\boldsymbol{x}|\boldsymbol{z})$, respectively, where ϕ and θ are the parameters of the encoder and the decoder. Additionally, it also assumes that the true posterior is an approximate Gaussian distribution with an approximately diagonal covariance.

To approximate the true posterior, the Kullback-Leibler (KL) divergence is minimized:

$$\begin{aligned} D_{KL}(q_\phi(\boldsymbol{z}|\boldsymbol{x})||p_\theta(\boldsymbol{z}|\boldsymbol{x})) &= \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log \frac{q_\phi(\boldsymbol{z}|\boldsymbol{x})}{p_\theta(\boldsymbol{z}|\boldsymbol{x})}\right] \\ &= \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log q_\phi(\boldsymbol{z}|\boldsymbol{x}) - \log p_\theta(\boldsymbol{x}|\boldsymbol{z}) - \log p_\theta(\boldsymbol{z})\right] + \log p_\theta(\boldsymbol{x}). \end{aligned} \tag{2.22}$$

Notice that $p_\theta(\boldsymbol{x})$ does not depend on $\boldsymbol{z}$ and thus can be taken out of the expectation. After rearrangement, we have

$$\log p_\theta(\boldsymbol{x}) - D_{KL}(q_\phi(\boldsymbol{z}|\boldsymbol{x})||p_\theta(\boldsymbol{z}|\boldsymbol{x})) = \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log p_\theta(\boldsymbol{x}|\boldsymbol{z})\right] - D_{KL}(q_\phi(\boldsymbol{z}|\boldsymbol{x})||p_\theta(\boldsymbol{z})). \tag{2.23}$$

This equation is the core equation for VAE (Doersch, 2016). Minimizing the KL divergence is equivalent to maximizing the right hand side, which is called the variational lower bound, i.e.,

$$\mathcal{L}_{VAE} = -D_{KL}(q_\phi(\boldsymbol{z}|\boldsymbol{x})||p_\theta(\boldsymbol{z})) + \mathbb{E}_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}\left[\log p_\theta(\boldsymbol{x}|\boldsymbol{z})\right]. \tag{2.24}$$

The first term is the regularization term which measures how far $q_\phi(\boldsymbol{z}|\boldsymbol{x})$ deviates from $p_\theta(\boldsymbol{z})$. The second term measures how likely the data is generated via the sampled latent variable $\boldsymbol{z}$, which typically is measured by a reconstruction error.

Directly optimizing the variational lower bound is difficult because the latent vector is a random variable. Thus a "reparameterization trick" (Kingma and Welling, 2013) can be used to solve the problem. Given the mean $\mu(\boldsymbol{x})$ and covariance matrix $\Sigma(\boldsymbol{x})$ for $q_\phi(\boldsymbol{z}|\boldsymbol{x})$, a variable z can be sampled via $z = \mu(\boldsymbol{x}) + \Sigma^{1/2}(\boldsymbol{x}) \times \epsilon$, where $\epsilon \sim \mathcal{N}(0, I)$. This approach essentially moves the sampling to an input layer. Therefore, given a data point and a sampled ϵ, the network is now deterministic and continuous. Hence we can use the chain rule to compute gradients and use gradient descent to update the parameters of the network.

2.3.4 Deep Learning in Image Registration

Finally, in this section, I will present three deep learning approaches for medical image registration. While there exist many deep learning-based image registration methods (Haskins et al., 2020), the presented three are most relevant to this dissertation.

Quicksilver. Quicksilver (Yang et al., 2017) is one of the first registration methods that uses deep learning. It is proposed to replace the costly numerical optimization for the LDDMM (Beg et al., 2005) by a deep regressor. Additionally, it parametrizes the deformation using the shooting formulation (Singh et al., 2013). The network takes a pair of patches from the moving and the target image at the corresponding location and outputs an initial vector-valued momentum patch. The ground-truth momentum is obtained via an optimization solution. The loss function is the 1-norm between the predicted momentum and the groundtruth. Diffeomorphism is a desirable property for registration. A transformation is diffeomorphism when it is smooth and bijective, and its inverse is also smooth. For fluid-based registration, diffeomorphsim is guaranteed when velocity field is sufficiently smooth (Dupuis et al., 1998). The benefit of using initial momentum as the parametrization of LDDMM is that it allows for the explicit control of the smoothness of the velocity and, therefore, the diffeomorphism: the velocity field is obtained by smoothing the predicted momentum.

VoxelMorph. VoxelMorph (Balakrishnan et al., 2018) is an unsupervised learning framework for image registration. The network aims to learn a mapping from a moving-target image pair to the deformation field that aligns them. The mapping is learned via a deep convolutional neural network

(CNN). The loss function of the network is an image similarity loss between the warped image and the target image, and a regularization loss on the deformation. It is the same as the energy equation (2.1) for traditional optimization-based methods, but this equation is now used to learn the network. Additionally, if the anatomical segmentations are available during training, the network can also learn to focus on the region of interest as the segmentation loss can be added as part of the loss function (Balakrishnan et al., 2019).

vSVF Mapping. An another deep learning based registration network is vSVF Mapping (Shen et al., 2019a). It also aims to replace the expensive numerical optimization via a deep network. The network is also learned in an unsupervised way, similar to VoxelMorph, but, instead of directly predicting the deformation, the network learns a vector momentum-parametrized stationary velocity field (vSVF) model from the training data. Equation 2.7 is used as the loss function. Similar to Quicksilver, the network predicts a momentum parametrization, and the deformation field is calculated via smoothing the momentum followed by integration of the velocity field.

CHAPTER 3: Efficient Registration of Images with Pathologies

This chapter presents a registration framework that aims to register images with pathologies. Inspired by the low-rank/sparse (LRS) framework (Liu et al., 2014, 2015), it is also an iterative registration and decomposition framework. Given an image with pathology, the framework decomposes an image with pathologies to a quasi-normal image and an abnormal part. The decomposition is based on principal component analysis (PCA). The quasi-normal image is then used to register the image to atlas space. As the decomposition and registration are interleaved with each other, both of them benefit when combined into an iterative framework. Contributions for the framework are as follows:

- *Normal* images, i.e. images from healthy patients, are used as the population. This is different from the original LRS framework which iteratively estimates quasi-normal images from a group of images with pathologies (interleaved with registration to a normal atlas). Instead, in the proposed framework, normal images are registered to the atlas only *once*. Additional registrations are performed only for the image with pathologies. Therefore, the computational cost is greatly reduced.

- When LRS is applied to a population of normal images and one image with pathologies, the most desirable decomposition would be to allocate all normal images to the low-rank part and to decompose only the image with pathologies into its low-rank and sparse components. While desirable, this will not happen in practice, because part of the normal images will also be allocated to the sparse part, causing image blurring. Therefore, in the proposed framework, the LRS decomposition is completely replaced. Specifically, the low-rank component is mimicked via a PCA basis obtained from the normal images in atlas space. The image with pathologies is decomposed into (i) a quasi-normal part which is *close* to the PCA space and

(ii) an abnormal part which has low total variation (TV) and replaces the sparse component in the LRS decomposition. This new decomposition is highly beneficial as it avoids image blurring (by only requiring *closeness* to the PCA space) and captures large pathologies (via TV) while avoiding attributing image detail and misalignment to the pathology as in LRS.

- The proposed framework is evaluated via a synthetic tumor dataset and BraTS2015 (Menze et al., 2014; Kistler et al., 2013). It shows similar performance to cost function masking while not requiring prior knowledge of the location of the pathology. Compared to LRS, it shows improved registration accuracy in both pathological regions and healthy region. Additionally, as it substantially reduces computational cost and requires much less memory than LRS, a GPU implementation is possible, resulting in much faster computation.

The remainder of this chapter is organized as follows: Section 3.1 reviews the LRS decomposition in detail and section 3.2 describes the proposed approach. Section 3.3 presents experimental results on a synthetic tumor dataset and BraTS2015, and section 3.4 concludes this chapter. The work presented in this chapter has been published in 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017) (Han et al., 2017)

3.1 Review of Low-Rank/Sparse (LRS)

An LRS decomposition aims at minimizing (Wright et al., 2009)

$$E(L, S) = \text{rank}(L) + \lambda \|S\|_0, \quad \text{s.t.} \quad D = L + S, \tag{3.1}$$

i.e., the goal is to find an additive decomposition of a data matrix $D = L + S$ such that L is low-rank and S is sparse. Here, $\|S\|_0$ denotes the number of non-zero elements in S and $\lambda > 0$ weighs the contribution of the sparse part, S, in relation to the low-rank part L. Neither rank nor sparsity is a convex function. Hence, to simplify the solution of this optimization problem, it is relaxed: the rank is replaced by the nuclear norm, and the sparsity term is replaced by the one-norm. As both of these norms are convex and $D = L + S$ is a linear constraint, the convex approximation to the LRS

decomposition can be obtained by minimizing the energy

$$E(L, S) = \|L\|_* + \lambda\|S\|_1, \quad \text{s.t.} \quad D = L + S, \tag{3.2}$$

where $\|\cdot\|_*$ is the nuclear norm. In imaging applications, D contains all the (vectorized) images: each image is represented as a column of D. The low-rank term captures common information across columns. The sparse term, on the other hand, captures uncommon/unusual information. As Equation (3.2) is convex, minimization results in a global minimum, e.g., computed via an augmented Lagrangian approach (Lin et al., 2010).

In practice, applying the LRS model requires forming the matrix D from all the images. D is of size $m \times n$, where m is the number of voxels, and n is the number of images. For 3D images, $m \gg n$ (typically). Assuming all images are spatially well-aligned, L captures the quasi-normal appearance of the images whereas S contains pathologies which are not shared across the images. However, in practice, the objective is image alignment and the images in D cannot be assumed to be aligned a-priori. Hence, the LRS decomposition steps can be alternated with image registration steps (Liu et al., 2014). Here the registrations are between all the low-rank images (which are assumed to be approximately pathology-free) and an atlas image. This approach is effective in practice, but can be computationally costly as it requires large amounts of memory. In particular, the matrix D has a large number of rows for typical 3D images, hence it can be costly to store. Furthermore, optimizing the LRS decomposition iteratively requires a singular value decomposition (SVD) at each iteration with a complexity of $\mathcal{O}(min\{mn^2, m^2n\})$ (Dongarra et al., 2018) for an $m \times n$ matrix. While large datasets are beneficial to capturing data variation, the quadratic complexity renders LRS computationally challenging in these situations. Additionally, since LRS does not consider spatial image information, small misalignments that is unavoidable in image registration, as well as image details, may be considered abnormal and allocated to the sparse part, causing blurring in estimated normal image parts.

3.2 Joint PCA/Image-Reconstruction model

The proposed model assumes that a collection of normal images is available. In fact, the goal is to register *one* pathological image to a normal-control atlas. Hence, all the normal images are first registered to the atlas using a standard image similarity measure. These normal images do not need to be re-registered during the iterative solution approach, resulting in a dramatic reduction of computational cost, which then allows using large image populations to capture normal data variation. Since *a priori* which images are normal is known, the low-rank part of LRS is mimicked by a PCA decomposition of the atlas-aligned normal images. Assuming PCA basis images $\{\beta_l\}$ and the mean image M are obtained, the model is now only concerned with a *single* image with pathologies I. Let $\hat{I}$ denote the image with pathologies after subtracting μ, i.e., $\hat{I} = I - \mu$, B the PCA basis matrix, and L and T are images of the same size[1] as I. The *first model* minimizes

$$E(T, \hat{L}, \boldsymbol{\alpha}) = \gamma \|\hat{L} - B\boldsymbol{\alpha}\|_1 + \|\nabla T\|_{2,1}, \quad \text{s.t.} \quad \hat{I} = \hat{L} + T, \tag{3.3}$$

akin to the TV-L1 model (Chan and Esedoglu, 2005), where $\|\nabla T\|_{2,1} = \sum_i \|\nabla T_i\|_2$ and i denotes spatial location. The *second model* minimizes

$$E(T, \hat{L}, \boldsymbol{\alpha}) = \frac{\gamma}{2} \|\hat{L} - B\boldsymbol{\alpha}\|_2^2 + \|\nabla T\|_{2,1}, \quad \text{s.t.} \quad \hat{I} = \hat{L} + T \tag{3.4}$$

and is akin to the Rudin-Osher-Fatemi (ROF) model (Rudin et al., 1992). Both models result in a TV term, T, which explains the parts of $\hat{I}$ which are (i) spatially contiguous, (ii) relatively large, and (iii) cannot be explained by the PCA basis, e.g., a tumor region. The quasi-low-rank part $\hat{L}$ remains close to the PCA space but retains fine image detail. Adding μ to $\hat{L}$ results in the reconstructed quasi-normal image L. In principle, model (3.3) would be preferred because of the attractive geometric scale-space properties of the TV-L1 model (Chan and Esedoglu, 2005). However, The model (3.4) is used in the experiments, as it is simpler to optimize and implement. Unfortunately,

[1] Images are vectorized; the spatial gradient ∇ is defined correspondingly.

just as the ROF model, it suffers from an intensity loss. This loss can be counteracted by adapting an iterative regularization approach (Osher et al., 2005) for the ROF model, which iteratively adds "noise" back to the original images. Specifically, equation (3.4) is first solved (obtaining $\tilde{L}_0 = \hat{L}$ and $\boldsymbol{\alpha}_0$), followed by a small number of regularization steps. For each iteration $k \geq 1$, the following equation is minimized:

$$E(T_k, \tilde{L}_k, \boldsymbol{\alpha}_k) = \frac{\gamma}{2}\|\tilde{L}_k - B\boldsymbol{\alpha}_k\|_2^2 + \|\nabla T_k\|_{2,1}, \quad \text{s.t.} \quad \hat{I}_k = \tilde{L}_k + T_k, \tag{3.5}$$

where $\hat{I}_k = \hat{I} + \tilde{L}_{k-1} - B\boldsymbol{\alpha}_{k-1}$. After N iterations, the TV part, T_N, will contain an approximation of the pathology, from which the quasi-normal image is obtained $\hat{L}_N = \hat{I} - T_N$. The quasi-normal image reconstructs pathological areas while retaining detailed image information in normal image areas.

3.3 Experiments

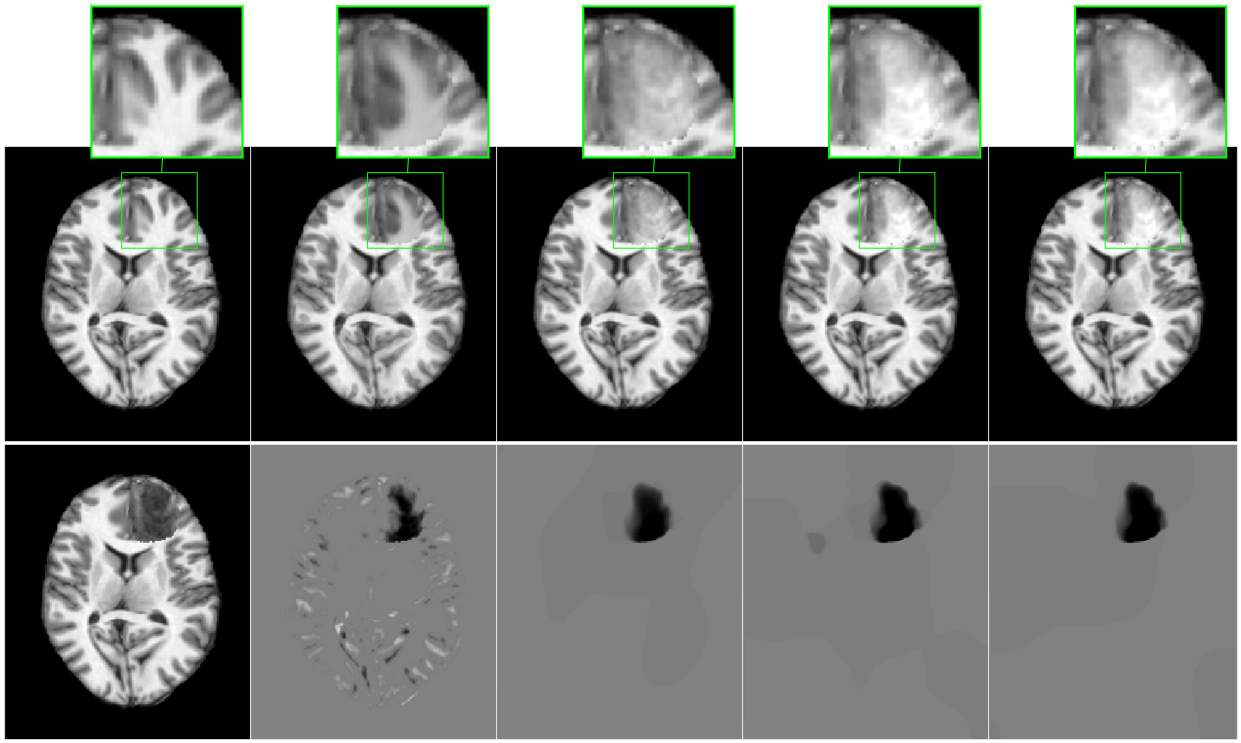
The ICBM 152 non-linear T1-weighted (T1w) atlas (2009a) (Fonov et al., 2009) is used as the normal control atlas. It is a $1 \times 1 \times 1$ mm magnetic resonance (MR) template with $197 \times 233 \times 189$ voxels. Images from the Open Access Series of Images Studies (OASIS)[2] (Marcus et al., 2007) are used to build the PCA model. The PCA model (3.4) is solved via a primal-dual hybrid gradient (PDHG) method (Goldstein et al., 2013). Images are aligned to the ICBM atlas via `NiftyReg` including affine (Ourselin et al., 2001; Modat et al., 2014) and b-spline registration (Rueckert et al., 1999; Modat et al., 2010). Localised normalized cross-correlation (LNCC) is used as the similarity measure in `NiftyReg`. For the 2D experiment, images are extracted from the same slice from 3D images.

3.3.1 Synthetic 2D Tumor Dataset

The performance of the model is first evaluated in 2D. Two hundred and fifty images from OASIS are selected as the population. Fifty distinct test cases are picked from the remaining set

[2]While some of patients are diagnosed with AD, the images do not show strong pathologies. Hence some of them are included in the normal population.

of OASIS images. These images are then registered to the BraTS 2015 T1w contrast-enhanced images (Menze et al., 2014; Kistler et al., 2013) with cost function masking, followed by pasting the BraTS tumors into these warped images as tumor segmentations are available. The registration is to simulate tumor mass effects. Each image is of size 197×233 with 1mm isotropic pixels. Fifty fixed normal images are selected as the population for LRS, to test a scenario which would still be computable in 3D given the high computational demand of LRS. Two hundreds and fifty normal images are selected for the proposed PCA model and the top 150 PCA modes are chosen as the PCA basis. The model is tested *without* regularization and *with* at most two regularization steps.



(a) (b) (c) (d) (e)

Figure 3.1: Example quasi-normal reconstructions. (a) ground truth (*top*) and tumor (*bottom*); (b)-(e) Reconstruction result (*top*) and tumor (*bottom*): (b) LRS; (c) PCA model w/o regularization; (d) PCA model w/ one and (e) w/ two regularization steps.

For each model, a cross-validation is performed. The 50 test cased is partitioned into 10 folds, with 9 folds for training and 1 fold for testing. Each model is trained with $\lambda = \{0.005, 0.0067, 0.0084, 0.01, 0.0117, 0.0133, 0.015\}$, for LRS, and $\gamma = \{0.5, 1, 1.5, 2, 2.5,3\}$, for PCA models. The atlas is registered to the ground-truth image. The resulting deformation is served as the

gold-standard registration. Then the atlas is also registered to reconstructed quasi-normal images by PCA models and by LRS. The *mean deformation error* is compared between each method. This is done in three areas: the tumor area, the normal areas near the tumor (within 10mm) and the normal areas far from the tumor (>10mm). The deformation errors in these areas are weighed by 4:1:1, for each model, to pick the parameter, i.e., λ or γ, that gives the smallest errors.

Figure 3.1 shows a good but blurry LRS reconstruction as the sparse part captures the tumor *and* misalignments. Also, the small and round left posterior ventricle in the ground truth image is not reconstructed faithfully by LRS. However, the PCA models capture only the tumor in T, resulting in a sharper and more precise reconstruction. Furthermore, regularization yields an even better tumor separation.

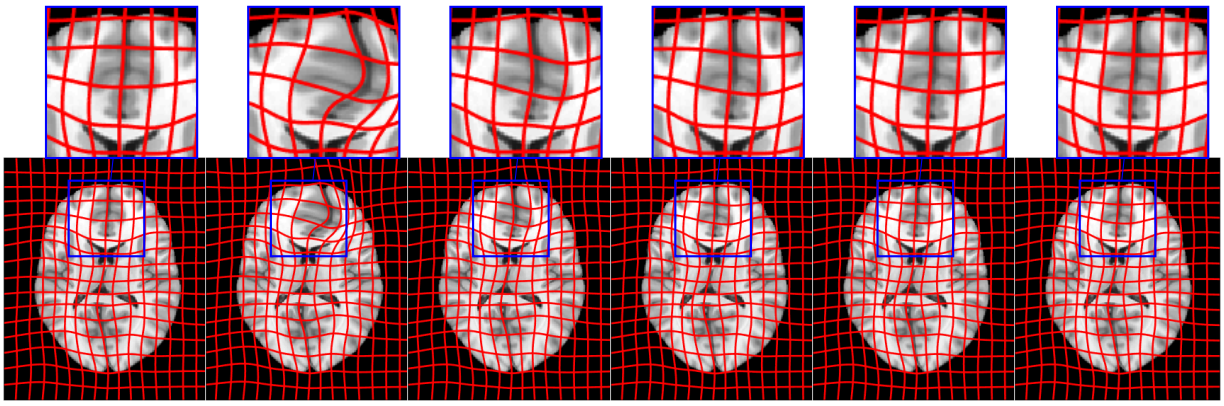


Figure 3.2: Example atlas-to-image registrations: (a) ground truth; (b) tumor; (c) LRS; (d) PCA model w/o regularization; (e) PCA model w/ one step and (f) w/ two regularization steps.

Figure 3.2 shows atlas-to-image registration results for images with and without tumor, LRS reconstruction and PCA-based reconstructions with and without regularization. Figure 3.3 shows the spatial error distributions, compared to the gold-standard registration. Errors are computed using Euclidean distance. Direct registration of the tumor image results in large deformation errors. Registration to the low-rank reconstruction greatly reduces the error in the tumor areas but retains errors near the cortex, mainly due to its blurry reconstruction. The PCA models further reduce deformation errors in the tumor areas *and* keep errors near the cortex low.

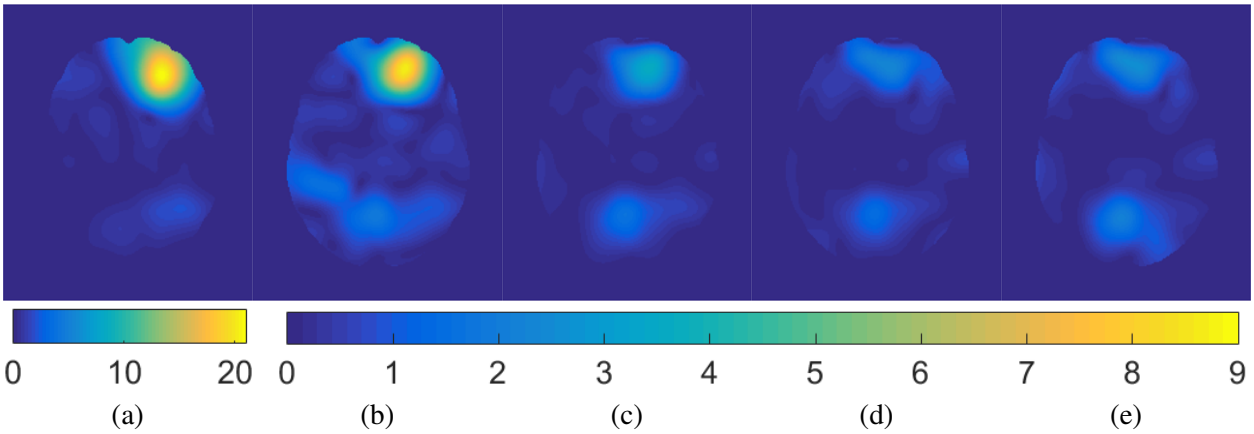


Figure 3.3: Example atlas-to-image registration errors [mm]: (a) tumor; (b) LRS; (c) PCA model w/o regularization; (d) PCA model w/ one step of regularization; (e) PCA model w/ two steps of regularization.

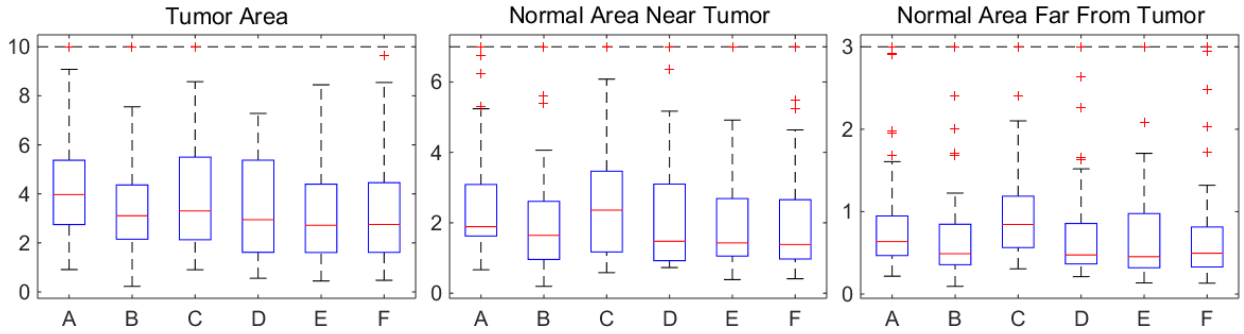


Figure 3.4: Mean deformation errors [mm] for test cases. *A:* Tumor image; *B:* cost function masking; *C:* LRS; *D:* PCA model w/o regularization; *E:* PCA w/ one and *F:* w/ two regularization steps.

Figure 3.4 shows mean deformation errors over all test cases in the 3 areas. Cost function masking is also added for comparison. Note that the tumors selected from BraTS to generate 2D test cases are relatively mild resulting in relatively small deformation errors even when using tumor images for registration. LRS (C) reduces errors in the tumor areas but has higher errors in the normal areas. PCA models (D, E, F) show better results in both the tumor and the normal areas. Paired t-tests between LRS and PCA models show statistically significant differences in all areas for the PCA models with regularization, and in the normal areas for the PCA model without regularization. Moreover, the PCA models with regularization show similar performance to cost function masking but do not require a tumor segmentation.

3.3.2 Synthetic 3D tumor Dataset

A 3D synthetic tumor dataset is also generated for evaluation. One hundred OASIS images are picked and the top fifty PCA modes as chosen the basis. Twenty test images are simulated with tumor (including mimicked mass effect). Each image is of size $197 \times 233 \times 189$. Different from the 2D experiment, the tumors for 3D test cases are picked randomly from BRATS, including cases with large tumors and deformations.

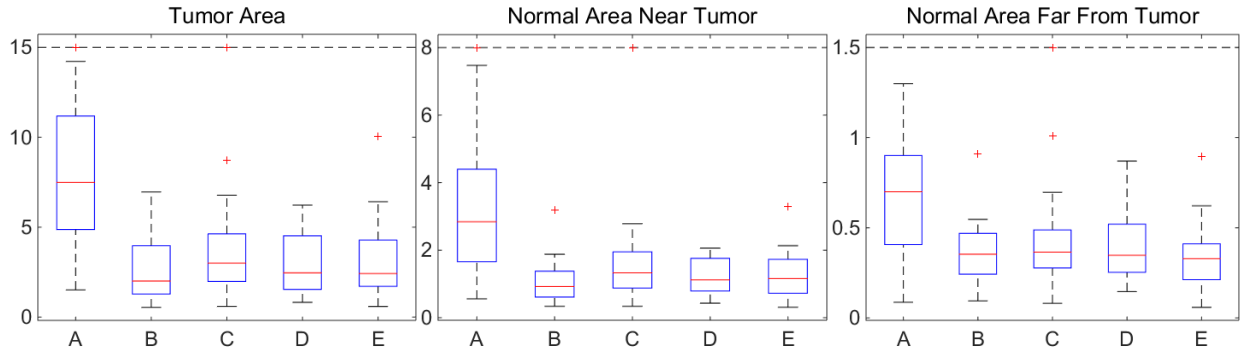


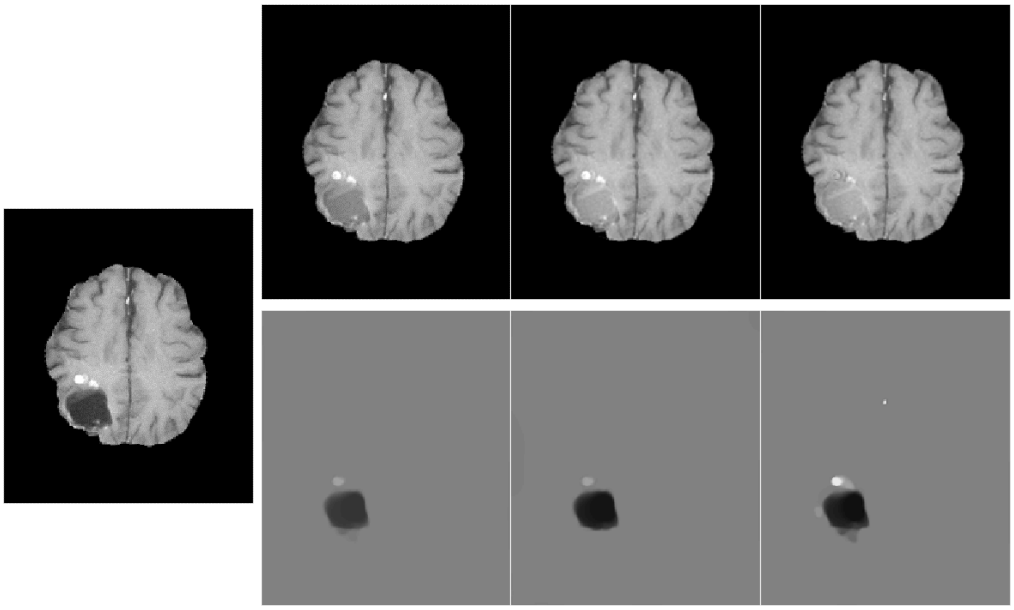
Figure 3.5: Mean deformation errors [mm] for 3D test cases. *A:* Tumor image; *B:* cost function masking; *C:* PCA model w/o regularization; *D:* PCA model w/ one and *E:* w/ two regularization steps.

For cross validation, twenty test cases are separated into ten 9:1 folds. The training parameters for PCA models are $\gamma = \{1, 1.5, 2, 2.5, 3\}$. Deformation errors in the three different areas are weighted as before, i.e., 4:1:1. Figure 3.5 shows box plots of the mean deformation errors. Directly registering to tumor images results in large errors. The quasi-normal images reconstructed by PCA models greatly reduce the deformation errors in all the areas. As in 2D, the PCA models show similar performance to cost function masking but do not require a tumor segmentation.

3.3.3 3D BraTS Dataset

Finally, the proposed PCA model is also applied to the Multimodal Brain Tumor Image Segmentation (BraTS) 2015 dataset (Menze et al., 2014; Kistler et al., 2013). As the BraTS data was acquired at different institutions and on different scanners, eighty BRATS T1w contrast-enhanced images are chosen as the population which show consistent image appearance and contain the full brain. To obtain the "normal population" for proposed PCA model, image intensities are locally imputed in the tumor areas, prior to computing the PCA basis, using the mean intensity over all

images that do not contain a tumor at that location. Also the first 50 PCA modes are picked as the basis.



(a) (b) (c) (d)

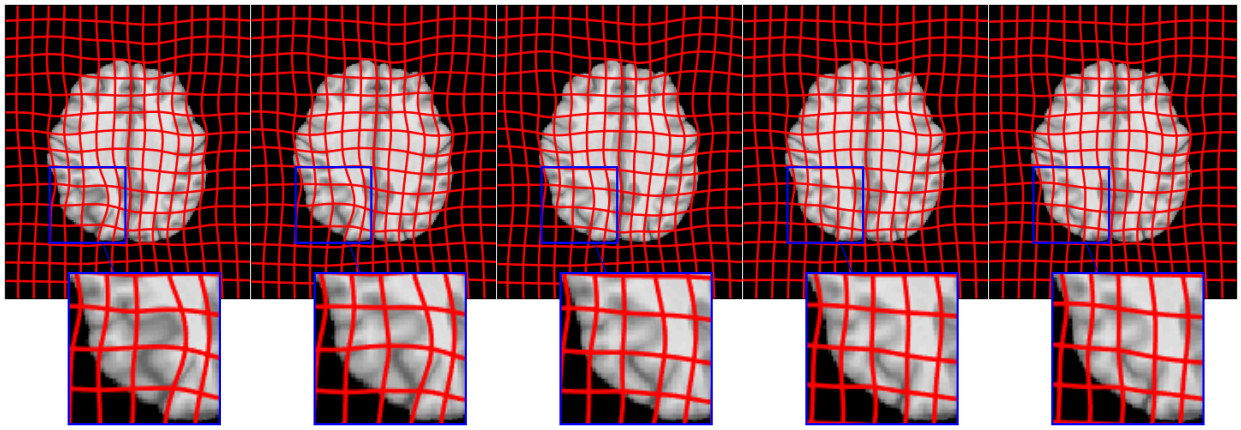
Figure 3.6: Example BRATS reconstructions: (a) tumor image; (b)-(d) reconstructions (*top*) and tumors (*bottom*); (b) PCA model w/o regularization; (c) PCA model w/ one and (d) w/ two regularization steps.

Figure 3.6 shows decomposition results for the PCA models. Parameter $\gamma = 5.0$ is selected for the model without and $\gamma = 2.0$ for models with regularization. The goal is to allocate as much of the tumor as possible to the abnormal part, T, while keeping the normal tissue in the quasi-normal part of the decomposition. Qualitatively, the proposed models identify tumor/normal areas, while retaining image details in normal tissue areas.

Finally, Figure 3.7 shows atlas-to-image registration results for the PCA models, the tumor image, and cost function masking. While the ground-truth registration result is not available, the results show the significant impact of the tumor on the registration, which is mitigated by cost function masking and proposed PCA models, in particular, with regularization.

3.3.4 Memory Use and Runtime

For LRS, $D \in \mathbb{R}^{m \times n}$, where m is the number of pixels/voxels and n the number of images. Each $197 \times 233 \times 189$ 3D image (stored as double) consumes about 65MB of memory. Hence,



(a) (b) (c) (d) (e)

Figure 3.7: Example BraTS atlas-to-image registration results: (a) tumor image; (b) cost function masking; (c) PCA model w/o regularization; (d) PCA model w/ one and (e) w/ two regularization steps.

3GB of memory is required to store D for $n = 50$. As the LRS algorithm (Lin et al., 2010) requires storing several variables of the size of D, memory use quickly becomes prohibitive, in particular for GPU implementations. The PCA model only stores *one* copy of the pre-computed PCA basis thereby substantially reducing memory use ($\approx$ 4GB/8GB for $n = 50$ in single/double precision) and consequentially facilitating larger sample sizes even on the GPU.

For the 3D cases, with $n = 50$, an LRS decomposition takes one hour to run and uses up to 40GB of memory thereby precluding a GPU implementation. Due to the low memory requirements of PCA models, a GPU implementation is possible resulting in a runtime of $\approx$3 minutes / decomposition. The 3D image registrations are computed on the CPU ($\approx$3 minutes). Therefore, with 6 registration iterations, proposed algorithm requires $\approx$40 minutes / test case and takes about 1 hour if extra regularization steps are computed, whereas the LRS approach takes $>$6 hours.

3.4 Conclusion

In this chapter, I proposed a PCA-based image decomposition model for an image with pathologies. A TV term is used capture large pathologies while the quasi-normal image is close to the PCA space. Compared to LRS, it avoids image blurring and retains image details in the quasi-normal image. Therefore, the registration using the PCA reconstructed image greatly improves the

results, especially in the normal region. Compared to cost function masking, the model reaches similar performance while not requiring prior tumor segmentations.

There are many ways in which the PCA models could be improved. For example, the decomposition approach is a compromise between model realism and model simplicity to allow for efficient computational solutions. While the total variation term succeeds at capturing the vast majority of large tumor masses and would likely work well for capturing volumes of resected tissue, the texture of pathological regions will not be appropriately captured and will remain in the quasi-normal image. Hence, it might be interesting to explore more realistic modeling assumptions to improve its quality. To obtain a more faithful quasi-normal image reconstruction would require more sophisticated modeling of the pathology. A natural approach could also be to perform this in the setting of a general adversarial network (Goodfellow et al., 2014) (GAN) to truly produce normal-looking quasi-normal images. As tumor images, for example, frequently exhibit mass effects, training and, formulating such a model could be highly interesting as one could attempt to model the expected mass effect as part of the GAN architecture.

Using a GAN, the pseudo-tumor dataset created for the experiments could be improved. As the tumor is simply copied on the warped OASIS images from the BraTS images (after the deformation simulation), tumor regions and non-tumor regions suffer from non-discontinuity and appearance differences. One way to create more realistic pseudo-tumors would be to use a GAN. The generator creates pseudo-tumor images to fool the discriminator while the discriminator tries to distinguish pseudo-tumor images from real tumor images. The discriminator could then also provide a good measure as to how realistic a pseudo-tumor is, for example, by computing what percentage of the pseudo-tumors could be detected.

The way that the PCA model is integrated into the decomposition could also be improved. Specifically, for computational simplicity, I only use the eigenspace created by a chosen number of PCA modes, but do not use the strength of these eigenmodes. This is a simple, yet reasonable strategy, to form a low-dimensional subspace capturing normal tissue appearance as long as a

pathology remains reasonably orthogonal to this subspace and hence would get assigned to the total variation part of the decomposition.

CHAPTER 4: Patient-Specific Registration of Pre-Operative and Post-Recurrence Brain Tumor Scans

This chapter presents a patient-specific registration framework that is to register pre-operative (pre) and post-recurrence (post) scans of patients with glioblastoma. This framework is an extension of the PCA based framework in Chapter 3 but is patient-specific, and it is specifically designed for this clinically relevant problem. In particular, it uses the post-scan, which is usually free from mass effect, to build a patient-specific PCA basis and registers the pre-scan to the patient space. This patient-specific PCA allows for more accurate modeling of the patient tissue and gives more reliable reconstruction, which in turn improves the registration results. Overall contributions are as follows:

- An automatic, repeatable, patient-specific registration framework is presented. The framework builds a patient-specific statistical model, which captures resections and mass effect reductions between pre-scans and post-scans within subject.
- The framework leverages the decomposition's TV term to intrinsically exclude the estimated pathology in case the image is not well-aligned to the target space.
- The proposed approach requires neither a complex tumor growth model nor manual interactions, such as tumor seeds or segmentations. It only requires a single modality and achieves significant accuracy improvements over six other methods.

The remainder of this chapter is organized as follows: Section 4.1 describes the proposed registration framework and section 4.2 presents qualitative and quantitative evaluations of the proposed approach, compared against other state-of-the-art methods. Section 4.3 summarizes the proposed framework. The work presented in this chapter has been published in 4th International

Brainlesion Workshop – Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries (BrainLes 2018) (Han et al., 2018a).

4.1 Patient-specific registration

In Chapter 3, I have introduced a joint PCA/image-reconstruction model (PCA-TV), which decomposes the pathological image into two parts: 1) normal tissue appearance is captured by a statistical (PCA) model; and 2) large pathologies are captured via a total-variation (TV) term, which avoids blurring of the normal tissue and retains fine details in the quasi-normal image. The reconstructed quasi-normal image is then used for atlas registration. To register the pre scan to the post scan, one could directly apply this method independently to each scan and then register the corresponding resulting quasi-normal images. However, this strategy would ignore the fact that these scans come from the same patient and the statistical model in the atlas space may not adequately capture the normal appearance for a specific patient; consequently, the registration quality may be impaired. In addition, the PCA-TV model registers quasi-normal images to the atlas during each iteration, but never uses the TV information. In case an image contains tumors with large mass effect, which is often the case for pre-scans, it is drastically misaligned with the population images. Hence, the decomposition may not work sufficiently, unless the image is well-aligned with the atlas. This is especially true during the first iteration of registration and decomposition. To overcome these shortcomings and improve the registration of *pre* and *post* scans, the following key adjustments are proposed to adapt the PCA-TV model to this specific problem.

4.1.1 PCA-TV-mask model

When the decomposition is computed in the first iteration, the image is only affinely aligned to the atlas. An affine transformation is far from accounting for the local change as well as the strong mass effect presented in the image. Therefore the decomposition results in many misalignments in the TV term and quasi-normal reconstruction is not always reliable. To migitate this issue, Otsu thresholding (Otsu, 1979) is applied to the TV image and a coarse mask of the pathological region (TV-mask) is obtained. This mask is then used during the registration, i.e., the quasi-normal image

is registered to the atlas but the TV-mask is used for cost-function masking of the tumor. Once the image is better aligned to the atlas via a deformable registration, the TV-mask is removed and the entire quasi-normal image is used for registration. This improvement is referred as the *PCA-TV-mask model*. This strategy is especially effective when a relatively small γ in equation (3.4) is set during the first iteration to restrict the part that is assigned to the TV term.

4.1.2 Patient-specific PCA

Considering i) that the *post* scan is relatively free from mass effects (e.g., except for scarring) and ii) that the tumor resection cavity is easily modeled via the TV term, the following three-step strategy is proposed. In the first step, the PCA-TV-mask model is applied to the *post* scan, resulting in a quasi-normal reconstructed image, in addition to registering the *post* scan to the atlas space. In the second step, the inverse transformation of the first step is used to map all normal images into the *post* scan space, and then construct a new PCA basis from this warped data. Importantly, now this new PCA basis together with the quasi-normal *post* image (now warped back to the patient space and used as atlas) can be used to run the PCA-TV-mask model on the *pre* scan, which is the final step of the proposed framework. Overall, this strategy allows *direct* registration between the *pre* and the *post* scans. Another advantage of using this patient-specific strategy is that by running PCA in the patient-specific space, the normal space spanned by the PCA basis is expected to be more consistent with the *pre* scan, which in turn improves the decomposition and registration results, when compared with the original PCA-TV framework.

4.2 Experiments

The proposed framework is evaluated on 10 pairs of *pre* and *post* clinically-acquired scans of patients diagnosed with de novo (primary) glioblastoma. Each timepoint contains native (T1) and T1-weighted contrast-enhanced (T1-CE), T2-weighted and FLAIR MRI. All modalities of each patient are skull-stripped, bias-field corrected, and affinely co-registered to the *pre* T1-CE scan of this patient that describes a $192 \times 256 \times 192$ volume with voxel size of $0.977 \times 0.977 \times 1.0[mm^3]$. For quantitative evaluation, manually seeded landmarks from two clinical experts are used. The

first expert placed 20 landmarks within 30[mm] from the tumor region and 30 landmarks outside the 30[mm] region in each *pre* scan. Then, both experts independently placed matching landmarks in the *post* scans. The landmarks placed by the first expert are considered the gold-standard and the ones placed by the second expert serve as a baseline comparison, referred to as `RATER`. In the experiments, only the T1 volumes are used from each patient and 6 iterations of registration and decomposition are run. The remaining 3 modalities were only used by the experts for seeding the landmarks. One hundred normal images from OASIS (Marcus et al., 2007) and 50 are selected as PCA basis. For registration, `NiftyReg` (Modat et al., 2010) is used as the registration method with the default settings and local normalized cross correlation as similarity measure (`--lncc 40`). The TV-mask is used in the first iteration when the image is only affinely aligned to the target image. After B-spline registration, the TV-mask is removed for subsequent iterations. Also the regularization steps is applied in the last three iterations. Parameter γ in equation (3.4) and equation (3.5) for the decomposition model is chosen as 1 if no regularization step is used and 2 if regularization steps are used. The following methods are compared against the proposed framework `AFFINE` (Modat et al., 2014), `GREEDY` (Avants et al., 2008), `DRAMMS` (Ou et al., 2011), `ANTS` (Avants et al., 2009), `NiftyReg` (Modat et al., 2010) and `PCA-TV` (Han et al., 2017). Although `PORTR` (Kwon et al., 2014) was specifically designed for this task and showed excellent results in the original work by using multiple image channels, it requires manual seeding on the tumor region and each tissue class, which makes it difficult to use. As the intent was to compare methods that do not require multiple modalities or manual interaction and hence more easily translate to clinical use, results from `PORTR` are omitted.

The mean landmark error is computed for each region of each patient (Figure 4.1) and note that all deformable methods are better than affine registration but worse than `RATER`. Compared to other deformable methods, the proposed patient-specific approach improves the results in the close-to-tumor region. Also results in the region far away from the tumor are improved, except when the proposed framework is compared with `NiftyReg` and the original `PCA-TV` model. In fact, as shown in Table 4.3, the improvements in the close-to-tumor region are statistically

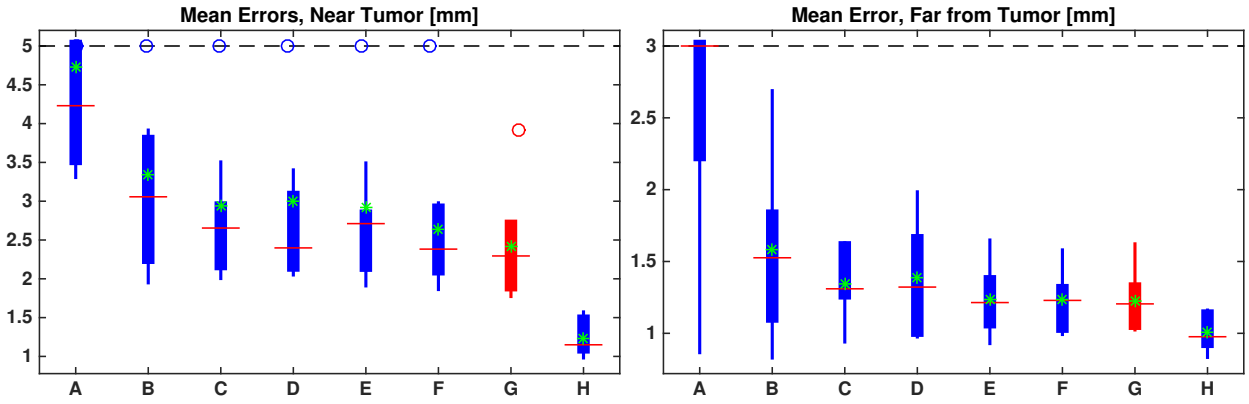


Figure 4.1: Boxplots of the mean landmark errors. For each method, the landmark errors are computed against the gold-standard. On each box, the red line is the median and the green star is the mean. The bottom and top edges of the box denote the 25th and 75th percentiles, respectively, the whiskers extend to the most extreme datas that are not considered outliers and the outliers are plotted in circle. (A) `AFFINE`; (B) `GREEDY`; (C) `DRAMMS`; (D) `ANTs`; (E) `NiftyReg`; (F) `PCA-TV`; (G) `PCA-PS`; (H) `RATER`. Result from proposed framework is plotted in red.

significant, assessed via a one-tailed paired Wilcoxon signed-rank test (Wilcoxon et al., 1970) with a Benjamini-Hochberg (Benjamini and Hochberg, 1995) procedure to control the false discovery rate at level $\alpha = 0.05$. For far-from-tumor regions, the results are only significant when compared to `AFFINE` and `GREEDY`. The effect sizes with each paired rank test are also calculated. Most of the tests result in large or medium effect sizes.

The statistics of the paired landmark errors are evaluated in both regions (Table 4.1). For each landmark, the differences of the errors between the proposed framework and competing methods are calculated. Compared to `RATER`, the proposed method shows worse performance on more than 50% of the landmarks. However, when comparing to other automatic registration methods, although at some landmarks the proposed method performs worse than others by less than 1.5[mm] near the tumor and 1[mm] far away from the tumor, as shown in the 5% statistics, it shows better performance on more than 50% of the landmarks. In fact, the improvement, especially near the tumor, can be larger than 5[mm], as shown in the 95% statistics in the table. Furthermore, on average, the proposed framework performs better than other registration methods by 0.5[mm] near the tumor and by less than 0.2[mm] far away from the tumor. This is consistent with the green stars

	Near Tumor[mm]						Far from Tumor[mm]					
	5%	25%	50%	75%	95%	Mean	5%	25%	50%	75%	95%	Mean
AFFINE	-1.03	0.22	1.65	3.71	7.58	2.32	-0.60	0.79	1.35	3.64	6.90	2.11
GREEDY	-1.01	-0.28	0.31	1.18	6.03	0.94	-0.80	-0.17	0.05	0.43	2.77	0.36
DRAMMS	-1.45	-0.55	0.18	0.79	4.68	0.52	-1.15	-0.28	0.14	0.52	1.16	0.13
ANTs	-1.44	-0.31	0.17	0.80	6.10	0.59	-0.68	-0.18	0.08	0.37	1.32	0.17
NiftyReg	-1.21	-0.19	0.12	0.60	3.35	0.51	-0.50	-0.14	-0.02	0.12	0.55	0.01
PCA_TV	-1.06	-0.29	0.11	0.57	2.08	0.23	-0.45	-0.14	0.00	0.15	0.55	0.01
RATER	-4.67	-1.86	-0.79	0.07	1.06	-1.18	-1.99	-0.74	-0.14	0.44	1.18	-0.21

Table 4.1: Statistic results for all paired landmark errors in both regions. For each landmark, the paired error is calculated; i.e., the landmark error of compared method is subtracted from the landmark error of the proposed method. This is to calculate the improvement obtained by the proposed method. For each compared method, the paired landmark errors are ranked and the statistics are shown in the table. The green boxes indicate results where errors from the proposed framework are smaller.

		AFFINE	GREEDY	DRAMMS	ANTs	NiftyReg	PCA-TV
p-values	Near	9.77e-4	4.90e-3	1.37e-2	1.86e-2	4.90e-3	3.22e-2
	Far	2.00e-3	1.37e-2	4.20e-2	0.116	0.423	0.385

Table 4.2: Significant tests between each method and proposed patient-specific framework

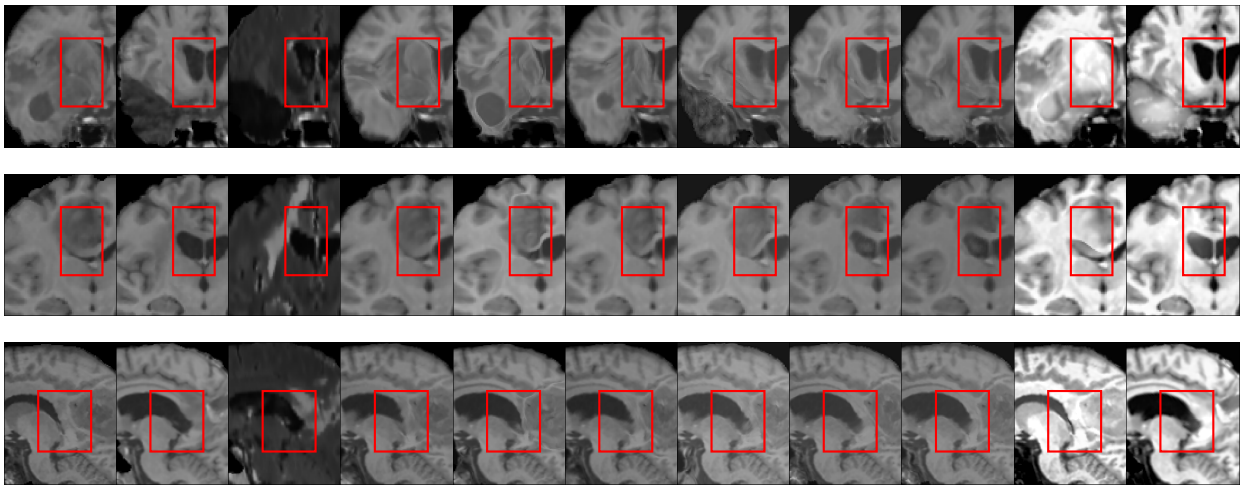
shown in Figure 4.1. The patient-specific method also improves over the PCA-TV model near the tumor which illustrates its utility and the benefit of the patient-specific model.

Finally, Figure 4.2 shows example results from three patients, where the *pre* scans are registered to the *post* scans. For the PCA-TV model and proposed patient-specific PCA-PS model, the quasi-normal images are reconstructed from each patient which are used to guide the registrations. Although the visual differences between the proposed method and the PCA-TV model are subtle, other results show that by modeling the pathologies registrations are qualitatively more accurate. Note that Figure 4.2(c) illustrates the T2-FLAIR scans for the *post* images, only for visualization

		AFFINE	GREEDY	DRAMMS	ANTs	NiftyReg	PCA-TV
p-values	Near	9.77e-4	4.90e-3	1.37e-2	1.86e-2	4.90e-3	3.22e-2
	Far	2.00e-3	1.37e-2	4.20e-2	0.116	0.423	0.385
effect sizes	Near	0.6268	0.5584	0.4900	0.4672	0.5584	0.4217
	Far	0.6040	0.4900	0.3989	0.2849	0.0570	0.0798

Table 4.3: p-values and effect sizes for one-tailed paired Wilcoxon signed-rank test. All methods (except for RATER) are compared with the proposed patient-specific framework. Green boxes indicate statistically significant results after false discovery rate correction or effect sizes that are at least medium (> 0.3).

purposes, to better depict the surgically-imposed cavities of these illustrated examples. All the applied registration methods use only the T1 volumes.



(a)pr-T1 (b)po-T1 (c)po-Fl (d) GRDY (e)DRMS (f)ANTs (g)NfRg (h)P-TV (i)P-PS (j)pr-qn (k)po-qn

Figure 4.2: Example registration results from three patients. (a) and (b) show the *pre* and the *post* T1 scans. (c) shows the *post* T2-FLAIR scans, only for visualization purposes. (d)-(i) show registration results of *pre* to *post* from GREEDY, DRAMMS, ANTs, NiftyReg, PCA-TV, and the proposed patient-specific model, PCA-PS. In addition, (j) and (k) show the quasi-normal reconstructions of the *pre* and *post* scans, respectively. The red box highlights major differences.

4.3 Conclusion

In this chapter, I proposed an automatic, repeatable, and accurate patient-specific registration framework based on a PCA-TV-mask model, which registers the pre-operative to the post-recurrence brain tumor MRI scan from the same patient. It does not require any manual interaction, neither in the form of segmentation nor as tumor seeding, and only requires a single modality. The validation results show that it is more effective than the PCA-TV model, as well as other registration methods that do not explicitly model pathologies.

The proposed patient-specific registration framework only requires a single modality, i.e., T1w images. While it simplifies the problem, pathologies that may not be obvious in the T1w image may get ignored. As a result, registration accuracy may be compromised. Four modalities are typically available: T1w, T2w, T1 contrast-enhanced, and FLAIR. All four modalities were used when experts identified the correspondence between the pre-operative image and the post-recurrence

image. This may be one of the reasons that, in the experiments, RATER always performs much better than other methods. Exploring formulations for different image sequences or modalities (or combination of modalities) would be interesting future work. However, it would require acquiring a normal population and run PCA for each modality.

The runtime of the algorithm is currently still in the order of 2-3 hours, as it requires two steps of PCA-TV-mask decompositions/registrations and one step to transform the population to the patient space and to run PCA. It could be substantially reduced by using a fast registration method, e.g., by combining the deep learning model presented later in Chapter 6. Furthermore, to speed-up the decompositions, one could explore numerical algorithms with faster convergence or reformulations of the decomposition itself.

Extending the framework to other longitudinal registration problems would also be an interesting direction, e.g., a patient-specific registration between the chronic and the acute image for patients with TBI.

CHAPTER 5: Brain Extraction from Images with and without Pathologies

In this chapter, I present a brain extraction[1] framework which can explicitly account for pathologies. Specifically, the model uses a three-part image decomposition: (1) normal tissue apperance is captured by principal component analysis (PCA), (2) pathologies are captured via a total variation term, and (3) the skull and surrounding tissue is captured by a sparsity term. Due to its convexity, the resulting decomposition model allows for efficient optimization. Additionally, decomposition and image registration steps are alternated to allow statistical modeling of normal tissue appearance in a fixed atlas coordinate system and, as a beneficial side effect, the decomposition model allows for the identification of potentially pathological areas and the reconstruction of a quasi-normal image in atlas space.

Contributions of this work are as follows:

- *(Robust) brain extraction:* The proposed method can reliably extract, with state-of-the-art performance, the brain from a wide variety of images, including images with normal appearance, slight and strong pathologies. Hence, it is a generic brain extraction approach.
- *Pathology identification:* The proposed model captures pathologies via a total variation term in the decomposition model.
- *Quasi-normal estimation:* The proposed model allows the reconstruction of a quasi-normal image, which has the appearance of a corresponding pathology-free or pathology-reduced image. This quasi-normal image also allows for accurate registrations to, e.g., a normal atlas.

[1] I avoid the commonly used term skull stripping, as my focus is to remove more than the skull from an image and to retain the parts of an image corresponding to the brain.

- *Extensive validation:* The proposed approach is extensively validated on four different datasets, two of which exhibit strong pathologies. The proposed method achieves state-of-the-art results on all these datasets using a *single* fixed parameter setting.

- *Open source:* The proposed method is available as open-source software.

The remainder of this chapter is organized as follows. Section 5.1 reviews related brain extraction approaches and Section 5.2 introduces the datasets that are used and discusses the proposed model, including the pre-processing, the decomposition and registration, and the post-processing procedures. Section 5.3 presents experimental results on 3D MRI datasets demonstrating that the proposed method consistently performs better than BET, BSE, ROBEX, BEaST, MASS and the deep learning approach for all four datasets. Section 5.4 concludes this chapter with a short discussion. The work presented in this chapter has been published in NeuroImage (Han et al., 2018b).

5.1 Review of Related Work

In this section, six widely-used brain extraction methods are reviewed, which cover a wide range of existing approaches. They are used in this work to compare to the proposed approach.

- *Brain Extraction Tool (BET):* BET (Smith, 2002) is part of the FMRIB Software Library (FSL) (Jenkinson et al., 2012) and is a widely used method for brain extraction. BET first finds a rough threshold based on the image intensity histogram, which is then used to estimate the center-of-gravity (COG) of the brain. Subsequently, BET extracts the brain boundary via a surface evolution approach, starting from a sphere centered at the estimated COG.

- *Brain Surface Extraction (BSE):* BSE (Shattuck et al., 2001) is part of BrainSuite (Shattuck and Leahy, 2002). BSE uses a sequence of low-level operations to isolate and classify brain tissue within T1-weighted MR images. Specifically, BSE uses a combination of diffusion filtering, edge detection and morphological operations to segment the brain. BrainSuite provides a user interface which allows for human interaction. Hence better performance may be obtained by interactive use of BSE.

- *Robust Learning-based Brain Extraction System (ROBEX):* ROBEX (Iglesias et al., 2011) is another widely used method which uses a random forest classifier as the discriminative model to detect the boundary between the brain and surrounding tissue. It then uses an active shape model to obtain a plausible result. While a modification of ROBEX for images with brain tumors has been proposed (Speier et al., 2011), its implementation is not available. Hence the standard ROBEX implementation is used for all experiments.

- *Deep Brain Extraction:* Additionally, a recently proposed deep learning approach for brain extraction (Kleesiek et al., 2016) is compared, which uses a 3D convolutional neural network (CNN) trained on normal images and images with mild pathologies. Specifically, it is trained on the IBSR v2.0[2] (Worth, 1996), LPBA40 (Shattuck et al., 2008) and OASIS (Marcus et al., 2007) datasets. This model is used as is without additional fine-tuning for other datasets.

- *Brain Extraction Based on non-local Segmentation Technique (BEaST):* BEaST (Eskildsen et al., 2012) is another recently proposed method, which is inspired by patch-based segmentation. In particular, it identifies brain patches by assessing candidate patches based on their sum-of-squared-difference (SSD) distance to known brain patches. BEaST allows using different image libraries to guide the brain extraction.

- *Multi-Atlas Skull Stripping (MASS):* MASS (Doshi et al., 2013), uses multi-atlas registration and label fusion for brain extraction. It has shown excellent performance on normal (IBSR, LPBA40) and close to normal (OASIS) image datasets. One of its main disadvantages is its runtime. An advantage of MASS, responsible for its performance and robustness, is that one can easily make use of dataset-specific brain templates. However, this requires obtaining such brain masks via costly manual segmentation. For a fair comparison to all other methods, and to test the performance of a given algorithm across a wide variety of datasets, templates provided along with the MASS software package are used for MASS's multi-atlas registration. These 15 anonymized templates are obtained from various studies.

[2]This is a different dataset than the IBSR dataset is used in this chapter.

In addition to these methods, many other approaches have been proposed. For example, A hybrid approach has been proposed (Ségonne et al., 2004) which combines watershed segmentation with a deformable surface model. Watershed segmentation is used to obtain an initial estimate of the brain region which is then refined via a surface evolution process. The 3dSkullStrip method is part of the AFNI (Analysis of Functional Neuro Images) package (Cox, 1996). It is a modified version of BET. In contrast to BET, it uses image data inside and outside the brain during the surface evolution to avoid segmenting the eyes and the ventricles.

Even though all these brain extraction methods exist and are regularly used, a number of challenges for automatic brain extraction remain:

- Many methods show varying performances on different datasets due to differences in image acquisition (e.g., slightly different sequences or differing voxel sizes). Hence, a method which can reliably extract the brain from images acquired with a variety of different imaging protocols would be desirable.

- Most methods only work for images which appear normal or show very minor pathologies. Strong pathologies, however, may induce strong brain deformations or strong localized changes in image appearance, which can impact brain extraction. For example, for methods based on registration, the accuracy of brain extraction will depend on the accuracy of the registration, which can be severely affected in the presence of pathologies. Hence, a brain extraction method which works reliably even in the presence of pathologies (such as brain tumors or traumatic brain injuries) would be desirable.

Therefore, Inspired by the low-rank + sparse (LRS) image registration framework (Liu et al., 2014) and previous work on image registration in the presence of pathologies in Chapter 3 (Han et al., 2017), I propose a brain extraction approach which can tolerate image pathologies (by explicitly modeling them) while retaining excellent brain extraction performance in the absence of pathologies.

5.2 Materials and Methods

5.2.1 Datasets

The ICBM 152 non-linear atlas (2009a) (Fonov et al., 2009) is used as the normal control atlas. ICBM 152 is a 1x1x1 mm template with 197×233×189 voxels, obtained from T1-weighted MRIs. Importantly, it also includes the brain mask. As the ICBM 152 atlas image itself contains the skull, a *brain-only* atlas is obtained simply by applying the provided brain mask.

Five different datasets are used for experiments. Specifically, one (OASIS, see below) of the datasets is used to build the PCA model and the remaining four to test the brain extraction approach.

OASIS. Images from the Open Access Series of Imaging Studies (OASIS) (Marcus et al., 2007) are used to build the PCA model for the proposed brain extraction approach. The OASIS cross-sectional MRI dataset consists of 416 sagittal T1-weighted MRI scans from subjects between 18 and 96 years of age. In this data corpus, 100 of the subjects over 60 years old have been diagnosed with very mild to mild Alzheimer's disease (AD). The original scans were obtained with in-plane resolution 1×1 mm (256×256), slice thickness = 1.25 mm and slice number = 128. For each subject, a gain-field corrected atlas-registered image and its corresponding masked image in which all non-brain voxels have been assigned an intensity of zero are available. Each image is resampled to $1 \times 1 \times 1$ mm isotropic voxels and is of size $176 \times 208 \times 176$.

Four datasets are used to evaluate the proposed approach, which all provide brain masks. Although in the study, I focus on T1-weighted images only, the proposed model can be applied to other modalities as long as the PCA model is also built from data acquired by the same modality. The datasets for validation are described below.

IBSR. The Internet Brain Segmentation Repository (IBSR) (Worth, 1996) contains MR images from 20 healthy subjects of age 29.1±4.8 years including their manual brain segmentations, provided by the Center for Morphometric Analysis at Massachusetts General Hospital. All coronal 3D T1-weighted spoiled gradient echo MRI scans were acquired using two different MR systems: ten scans (4 males and 6 females) were performed on a 1.5T Siemens Magnetom MR system (with in-plane resolution of 1×1 mm and slice thickness of 3.1 mm); another ten scans (6 males and 4

females) were acquired from a 1.5T General Electric Signa MR system (with in-plane resolution of 1×1 mm and slice thickness of 3 mm).

LPBA40. The LONI Probabilistic Brain Atlas (LPBA40) dataset of the Laboratory of Neuro Imaging (LONI) (Shattuck et al., 2008) consists of 40 normal human brain volumes. LPBA40 contains images of 20 males and 20 females of age 29.20 ± 6.30 years. Coronal T1-weighted images with slice thickness 1.5 mm were acquired using a 1.5T GE system. Images for 38 of the subjects have in-plane resolution of 0.86×0.86 mm; the images for the remaining two subjects have a resolution of 0.78×0.78 mm. A manually segmented brain mask is available for each image.

BraTS: Twenty T1-weighted image volumes of low and high grade glioma patients from the Brain Tumor Segmentation (BraTS) dataset are used (Menze et al., 2014) that include cases with large tumors, deformations, or resection cavities. The images that are used are not available as part of the BraTS challenge as these have already been pre-processed (i.e., brain-extracted and co-registered). Instead, a subset of twenty of the originally acquired images is obtained. The BraTS dataset is challenging as the images were acquired with different clinical protocols and various different scanners from multiple ($n = 19$) institutions (Bakas et al., 2017). The subset of twenty images is from six different institutions. Furthermore, the BraTS images have comparatively low resolution and some of them contain as few as 25 axial slices (with slice thickness as large as 7mm). The in-plane resolutions vary from 0.47×0.47 mm to 0.94×0.94 mm with image grid sizes between 256×256 and 512×512 pixels. The brain is manually segmented in these images to obtain an accurate brain mask for validation.

TBI. Finally, a Traumatic Brain Injury (TBI) dataset is used which contains 8 TBI images as well as manual brain segmentations. These are standard MPRAGE (Brant-Zawadzki et al., 1992) T1-weighted images with no contrast enhancement. They have been resampled to 1×1×1 mm isotropic voxel size with image size between $192 \times 228 \times 170$ and $256 \times 256 \times 176$. Segmentations are available for healthy brain, hemorrhage, edema and necrosis. To generate the brain masks, the union of healthy tissue and necrosis is always used. Also included are hemorrhage and edema if they are contained within healthy brain tissue.

Figure **??** shows example images from each dataset to illustrate image variability. IBSR and LPBA40 contain images from normal subjects and include large portions of the neck; BraTS has very low out-of-plane resolution; and the TBI dataset contains large pathologies and abnormal skulls.

5.2.2 Dataset processing

5.2.2.1 PCA model

One hundred images and their brain masks are randomly picked to build the PCA model of the brain. Specifically, the brain-masked images are registered to the brain-masked ICBM atlas using a B-spline registration. `NiftyReg` (Modat et al., 2010) is used to perform the B-spline registration with local normalized cross-correlation (LNCC) as similarity measure. To normalize image intensities, an affine transform is applied to the image intensities of the warped images so that the 1st percentile is mapped to 0.01 and 99th percentile is mapped to 0.99 and then image intensities are clamped to be within $[0, 1]$. PCA is then performed on the then registered and normalized images and the top 50 PCA modes are retained, which preserve 63% of the variance, for the statistical appearance model. This is similar to an active appearance model (Cootes et al., 2001).

5.2.2.2 IBSR refined segmentation

For IBSR, segmentations of the brain images into white matter, gray matter and cerebrospinal fluid (CSF) are provided. While, in principle, the union of the segmentations of white matter, gray matter and CSF should represent the desired brain mask, this is not exactly the case (see Figure 5.2). To alleviate this issue for each segmentation, morphological closing is used to fill in remaining gaps and holes inside the brain mask and, in particular, to disconnect the background inside the brain mask from the surrounding image background. The structuring element for closing is a voxel and its 18 neighborhood[3]. Then the connected component for the background is found and its complement is considered the brain mask. Figure 5.2 shows the pre-processing result after these refinement

[3]The 18-voxel connectivity is also used for other morphological operations in this chapter.

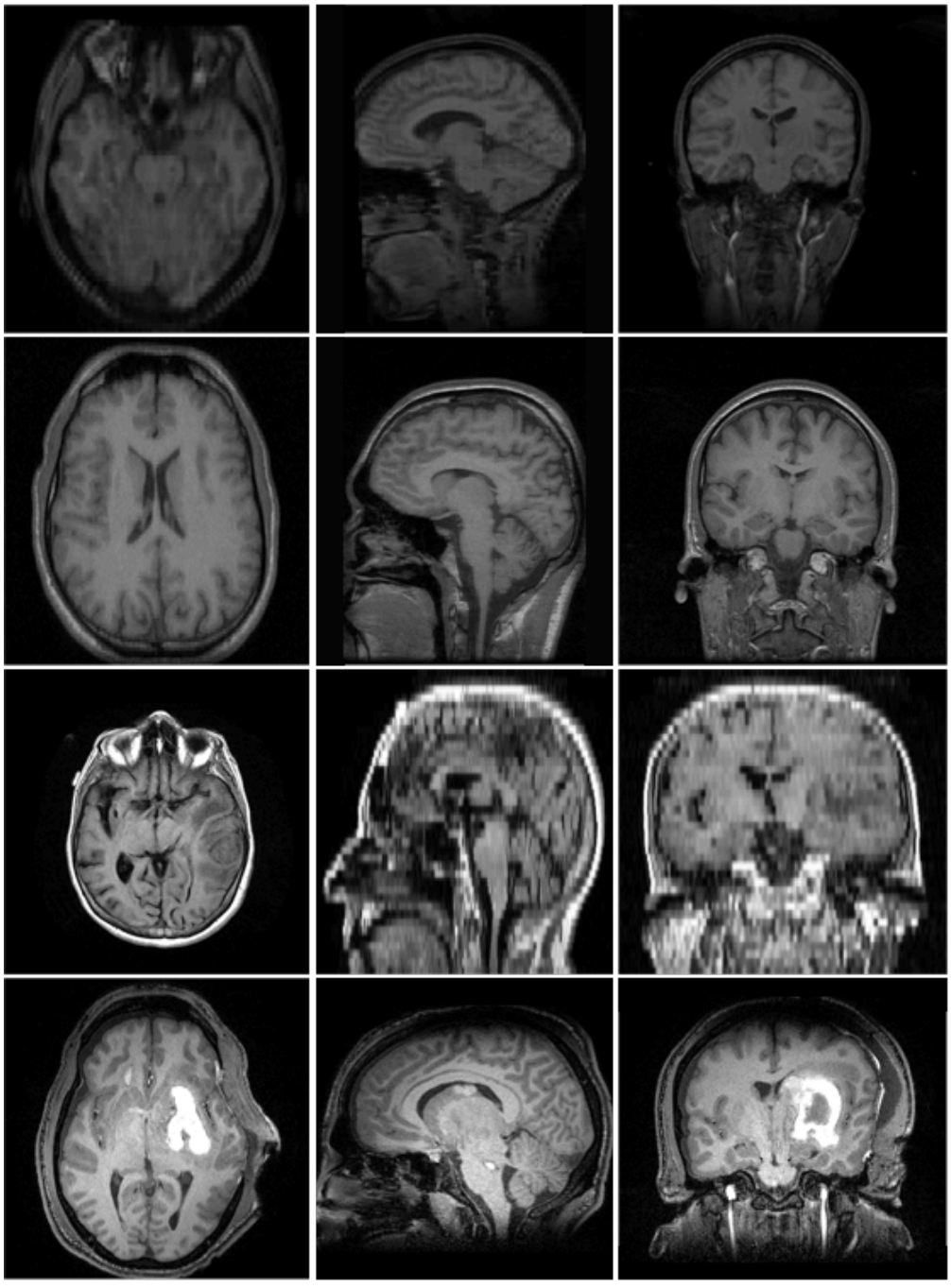


Figure 5.1: Illustration of image appearance variability on a selection of images from each (evaluation) database. From *top* to *bottom*: IBSR, LPBA40, BraTS and TBI.

steps, compared to the original IBSR segmentation (i.e., the union of white matter, gray matter, and the CSF).

(a) (b) (c)

Figure 5.2: Example coronal slice of (a) an IBSR MR brain image, (b) the corresponding original IBSR brain segmentation (i.e., union of white matter, gray matter and CSF) and (c) the refined brain segmentation result.

5.2.3 Proposed brain extraction approach

Recall in Chapter 3, a joint PCA/image-reconstruction model (PCA-TV) was proposed for improved and efficient registration of images with pathologies. In the PCA-TV model, an image with pathologies is decomposed into two parts: a quasi-normal image, which is pathology-reduced, and a pathological part (using a total variation term). It assumes that the available images have already been brain-extracted. However, the total variation term cannot reliably remove non-brain parts of the image. This shortcoming of the previous strategy motivated the new approach for brain extraction. Specifically, the new approach addresses these issues via an explicit decomposition into three instead of two components, introduces spatially-dependent weights, and provides an overall strategy customized for brain extraction.

The following sections describe how the proposed brain extraction approach builds upon the principles of the PCA-TV model (Section 5.2.3.1), and discusses image pre-processing (Section 5.2.3.2), the overall registration framework (Section 5.2.3.3), and post-processing steps (Section 5.2.3.4).

5.2.3.1 Joint PCA-Sparse-TV model

The PCA-TV model captures the pathological information well, but it does not model non-brain regions (such as the skull) appropriately. The skull is, for example, usually a thin, shell-shape structure and other non-brain tissue may be irregularly shaped with various intensities. The only commonality is that all these structures surround the brain. Specifically, if a test image is aligned to the atlas well, these non-brain tissues should *all* be located outside the atlas' brain mask. Hence, these non-brain regions are rejected via a spatially distributed sparse term. The sparsity is penalized heavily inside the brain and relatively little on the outside of the brain. This has the effect that it is very cheap to assign voxels outside the brain to the sparse term; hence, these are implicitly declared as brain outliers. Of course, if I would already have a reliable brain mask I would not need to go through any modeling. Instead, I assume that the initial affine registration provides a good *initial alignment* of the image, but that it will be inaccurate at the boundaries. Therefore a constant penalty is added close to the boundary of the atlas brain mask. Specifically, two masks are created: a two-voxel-eroded brain mask, which is confident to be within the brain and a one-voxel-dilated brain mask, which is confident to include the entire brain. The following model is then obtained:

$$E(S,T,\hat{L},\boldsymbol{\alpha}) = \frac{1}{2}\|\hat{L} - B\boldsymbol{\alpha}\|_2^2 + \gamma\|\nabla T\|_{2,1} + \|\boldsymbol{\Lambda} \odot S\|_1, \quad \text{s.t.} \quad \hat{I} = \hat{L} + S + T \tag{5.1}$$

where $\boldsymbol{\Lambda} = \Lambda(\boldsymbol{x}) \geq 0$ is a spatially varying weight

$$\Lambda(\boldsymbol{x}) = \begin{cases} \infty, & \boldsymbol{x} \in \text{Eroded Mask (inside)} \\ \lambda, & \boldsymbol{x} \in \text{Dilated Mask and } \boldsymbol{x} \notin \text{Eroded Mask (at boundary)} \\ 0, & \boldsymbol{x} \notin \text{Dilated Mask (outside)} \end{cases} \tag{5.2}$$

with $\boldsymbol{x}$ denoting the spatial location. Further, in Equation (5.1), $\odot$ indicates an element-wise product and $\gamma \geq 0$ weighs the total variation term.

This model is referred as the joint PCA-Sparse-TV model. It decomposes the image into three parts. Similar to the PCA-TV model, the quasi-low-rank part $\hat{L}$ remains close to the PCA space and the TV term, T, captures pathological regions. Here, the PCA basis is generated from normal images that have been already brain-extracted. Therefore $\hat{L}$ only contains the brain tissue. Different from the previous model, a spatially distributed sparse term, S, is added which captures tissue outside the brain, e.g., the skull. In effect, since Λ is very large inside the eroded mask, none of the image inside the eroded mask will be assigned to the sparse part. Conversely, all of the image outside the dilated mask will be assigned to the sparse part. This PCA-Sparse-TV model is then integrated into the low-rank registration framework. It includes three parts: pre-processing, iterative registration and decomposition, and post-processing which will be discussed in the following.

5.2.3.2 Pre-processing

Intensity normalization. Given a test image from in which the brain is to be extracted, The image intensities is first affinely transformed to a standardized range of $[0, 1000]$. Note that the PCA model of section 5.2.2.1 is built based on images with intensities standardized to $[0, 1]$. The different standardization is necessary here as the bias field correction algorithm (Sled et al., 1998) that is used later removes negative and small intensity values (< 1) followed by a log transform of the intensities. Specifically, the 1st and the 99th percentile of the voxel intensities are first computed. Then the image intensities of the entire image are affinely transformed such that the intensity of the 1st percentile is mapped to 100 and of the 99th percentile to 900. As this may result in intensities smaller than zero or larger than 1000 for the extreme ends of the intensity distribution, the intensities are clamped to be within $[0, 1000]$.

Atlas registration. Next, the intensity-normalized input image is first aligned to the non brain-extracted atlas. Then, the result from the first step is affinely registered to the brain-extracted atlas, but this time a one-voxel-dilated brain mask in atlas space is used; this step has the effect of ignoring parts of the image which are not close to the brain in the registration and it gives a better alignment in the brain region. For both steps `reg_aladin` of `NiftyReg` (Modat et al., 2014) is used disabling symmetric registration (`-noSym`). The first registration initializes the transformation

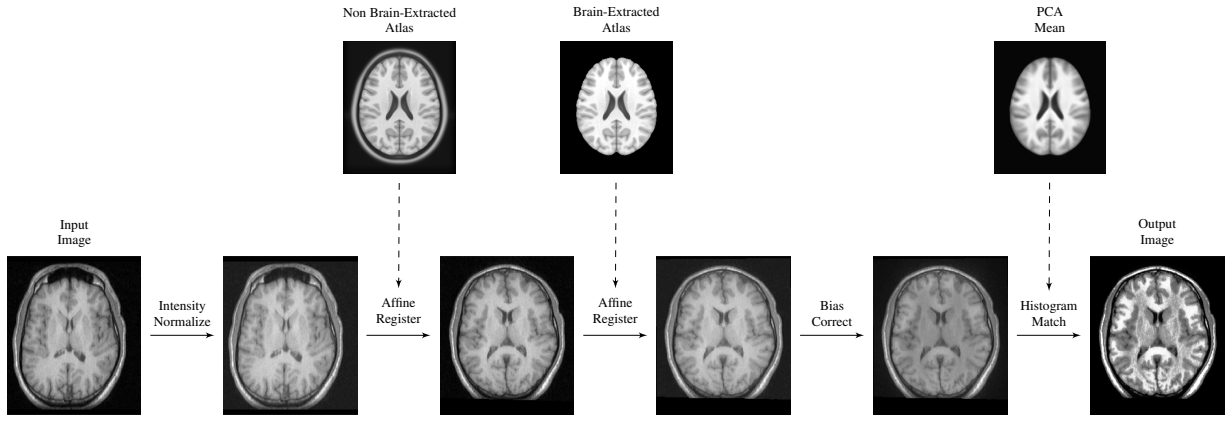


Figure 5.3: Preprocessing flow chart: Input image is the original image. Eventually, the output image will be fed into the registration/decomposition framework.

using the center of gravity (CoG) of the image. Note that the differing intensity range of the atlas and the image is immaterial in this step as the registration uses local normalized cross-correlation as the similarity measure.

Bias field correction. Next, N4ITK (Tustison et al., 2010) is used, a variant of the popular non-parametric non-uniform intensity normalization (N3) algorithm (Sled et al., 1998), to perform bias field correction. As the image has been affinely aligned to the atlas in the previous step, the two-voxel-eroded brain mask is used as the region for bias field estimation. Specifically, the `N4BiasFieldCorrection` function in `SimpleITK` (Lowekamp et al., 2013) is used with its default settings.

Histogram matching: The final step of the pre-processing is histogram matching. The histogram of the bias corrected image is matched with the histogram of the mean image of the population data only within the two-voxel-eroded brain mask. Then histogram matching is applied to the bias corrected image. This histogram matched image is then the starting point for the brain extraction algorithm and it is now in an intensity range comparable to the PCA model.

Figure 5.3 shows a flowchart of the proposed pre-processing approach.

5.2.3.3 Registration framework

Similar to the PCA-TV model, *image decomposition* steps using the PCA-Sparse-TV model and *registration to the brain-extracted atlas* are alternated. Six iterations are used in the brain

extraction framework. In the first iteration ($k = 1$), the images are in the original space. The input image $I_1 = I$ is decomposed into the quasi-normal ($L_1 = \hat{L}_1 + M$), sparse (S_1), and total variation (T_1) images by minimizing the energy from Equation (5.1). A pathology-free or pathology-reduced image, R_1, is then obtained by adding the sparse and the quasi-normal images of the decomposition: $R_1 = L_1 + S_1$.

For the next two iterations ($k = \{2, 3\}$), the affine transform Φ_k^{-1} is obtained by affinely registering the pathology-reduced images from the previous iteration, R_{k-1} (i.e., $R_{k-1} = L_{k-1} + S_{k-1}$), to the brain-extracted atlas[4]. The one-voxel-dilated brain mask is used for cost-function masking which allows the registration to focus only on the brain tissue. This is important as the first few registrations will not be very precise since they are only based on an affine deformation model. The main objective is to reduce the pathology within the brain. Only after these initial steps, when a good initial alignment has already been obtained, the quasi-normal image (excluding the non-brain regions) is used to perform the registration. The transform Φ_k^{-1} is then applied to transform the previous input images to atlas space and a new input image is obtained, I_k, (i.e., $I_k = I_{k-1} \circ \Phi_k^{-1}$). Equation (5.1) is minimized again to obtain new decomposition results (L_k, S_k, T_k). These decomposition/affine-registration steps are repeated two times, which is empirically determined to be sufficient for convergence. These affine registration steps result in a substantially improved alignment in comparison to the initial affine registration by itself.

The last three iterations ($k = \{4, 5, 6\}$) repeat the same process, but are different in the following aspects: (i) now a B-spline registration is used instead of the affine registration; (ii) the pathology-reduced image and cost function masking are only used for the first B-spline registration step as in the previous affine steps. For the remaining two steps, the quasi-normal images are used $L_{k:k=\{5,6\}}$ as the moving images. The use of the mask is no longer necessary as registrations are now performed using the quasi-normal image; (iii) the non-greedy registration strategy of the original low-rank + sparse framework (Liu et al., 2015) is used, in which the quasi-normal image

[4]The standard image-registration notation is followed. I.e., a map Φ^{-1} is defined in the space an image is deformed to. Thus this is the space of the atlas image. Conversely, Φ maps an image from the atlas space back into the original image space and hence is defined in the original image coordinate space.

is deformed back to the image space of the third iteration (after the affine steps) in order to avoid accumulating deformation errors.

These steps further refine the alignment, in particular, close to the boundary of the brain mask. After the last iteration, the image is well-aligned to the atlas and all the transforms are obtained from the original image space to atlas space. As a side effect, the algorithm also results in a quasi-normal reconstruction of the image, L_6, an estimate of the pathology, T_6, and an image of the non-brain tissue S_6, all in atlas space.

Algorithm 1: Algorithm for Brain Extraction

Input : Image I, Brain-Extracted Atlas A, Atlas Mask A_M
Output : Brain-Extracted Image I_B and mask I_M

I_1, Φ_1^{-1} = pre-processing(I);
for $k \leftarrow 1$ **to** 6 **do**
  if $k \geq 2$ **then**
    if $k \leq 3$ **then**
      find Φ_k^{-1}, s.t., $R_{k-1} \circ \Phi_k^{-1} = A$ and Φ_k^{-1} is affine;
    else if $k == 4$ **then**
      find Φ_k^{-1}, s.t., $R_{k-1} \circ \Phi_k^{-1} = A$ and Φ_k^{-1} is B-spline;
    else
      find Φ_k^{-1}, s.t., $(L_{k-1} \circ \Phi_{k-1}) \circ \Phi_k^{-1} = A$ and Φ_k^{-1} is B-spline;
    end
    $I_k = I_{k-1} \circ \Phi_k^{-1}$;
  Decompose I_k, s.t., $I_k = L_k + S_k + T_k$;
  if $k \leq 3$ **then**
    $R_k = L_k + S_k$;
  end
end
I_B, I_M = post-processing(A_M, $\{\Phi_k^{-1}\}$).

5.2.3.4 Post-processing

Post-processing consists of applying to the atlas mask the inverse transforms of the affine registrations in the pre-processing step and the inverse transforms of the registrations generated in the framework described in section 5.2.3.3. The warped-back atlas mask is the brain mask for the original image. To extract the brain in the original image space, the brain mask is simply applied on the original input image. All subsequent validations are performed in the original image space.

Algorithm 1 summarizes these steps as pseudo-code.

5.3 Experimental results

The following experiments are for brain-extraction from T1-weighted MR images. Note that the proposed approach can be easily adapted to images from other modalities, as long as the atlas image and the images from which the PCA basis is computed are from the same modality.

5.3.1 Experimental setup

The proposed method is evaluated on all four evaluation datasets. For comparison, the performance of BET, BSE, ROBEX, BEaST, MASS and CNN on these datasets are also assessed. Specifically, BET v2.1 as part of FSL 5.0, BSE v.17a from BrainSuite, ROBEX v1.2, BEaST (`mincbeast`) v1.90.00, and MASS v1.1.0 are used. The PCA model is solved via a primal-dual hybrid gradient method (Goldstein et al., 2013). In addition, the decomposition is implemented on the GPU and the algorithm is ran on an NVIDIA Titan X GPU (Nickolls et al., 2008; Givon et al., 2015).

5.3.2 Evaluation Measures

All brain extraction approaches are evaluated using the measures listed below.

Dice coefficient Given two sets X and Y (containing the spatial voxel positions of a segmentation), the Dice coefficient $D(X, Y)$ is defined as

$$D(X, Y) = \frac{2|X \cap Y|}{|X| + |Y|}, \tag{5.3}$$

where $X \cap Y$ denotes set intersection between X and Y and $|X|$ denotes the cardinality of set X.

Average, maximum and 95% surface distance Symmetric surface distances are also used to measure between the automatic brain segmentation and the gold-standard brain segmentation. The distance of a point x to a set of points (or set of points of a triangulated surface S_A) is defined as

$$d(x, S_A) = \min_{y \in S_A} d(x, y), \tag{5.4}$$

where $d(x, y)$ is the Euclidean distance between the point x and y. The average symmetric surface distances between two surfaces S_A and S_B is then defined as

$$ASD(S_A, S_B) = \frac{1}{|S_A| + |S_B|} \times (\sum_{x \in S_A} d(x, S_B) + \sum_{y \in S_B} d(y, S_A)), \tag{5.5}$$

where $|S_A|$ denotes the cardinality of S_A (Yeghiazaryan and Voiculescu, 2015) (i.e., number of elements if represented as a set or surface area if represented in the continuum). To assess behavior at the extremes, the maximum symmetric surface distance is also reported, as well as the 95th percentile symmetric surface distance which is less prone to outliers. These are defined in analogy, i.e., by computing all distances from surface S_A to S_B and vice versa followed by the computation of the maximum and the 95th percentile of these distances.

Sensitivity and specificity Sensitivity, i.e., true positive (TP) rate and specificity, i.e., true negative (TN) rate are also measured. Here TP denotes the brain voxels which are correctly labeled as brain; TN denotes the non-brain voxels correctly labeled as such. Furthermore, the false negatives (FN) are the brain voxels incorrectly labeled as non-brain and the false positives (FP) are the non-brain voxels which are incorrectly labeled as brain. Let V be the set of all voxels of an image, and X and Y the automatic brain segmentation and gold-standard brain segmentation, respectively. The sensitivity and specificity are then defined as follows (Sonka and Fitzpatrick, 2000) :

$$sensitivity = \frac{TP}{TP + FN} = \frac{|X \cap Y|}{|Y|} \tag{5.6}$$

$$specificity = \frac{TN}{TN + FP} = \frac{|V| - |X \cup Y|}{|V| - |Y|} \tag{5.7}$$

5.3.3 Datasets of Normal images: IBSR/LPBA40

IBSR results: Figure 5.4 shows the box-plots summarizing the results for the IBSR dataset. BEaST does not work well when applied directly on the IBSR images. This is due to failures with the initial spatial normalization (in 5 cases the computations themselves fail and in 10 cases the results are poor). Therefore, in the experiment, the same affine registration to atlas space is applied as

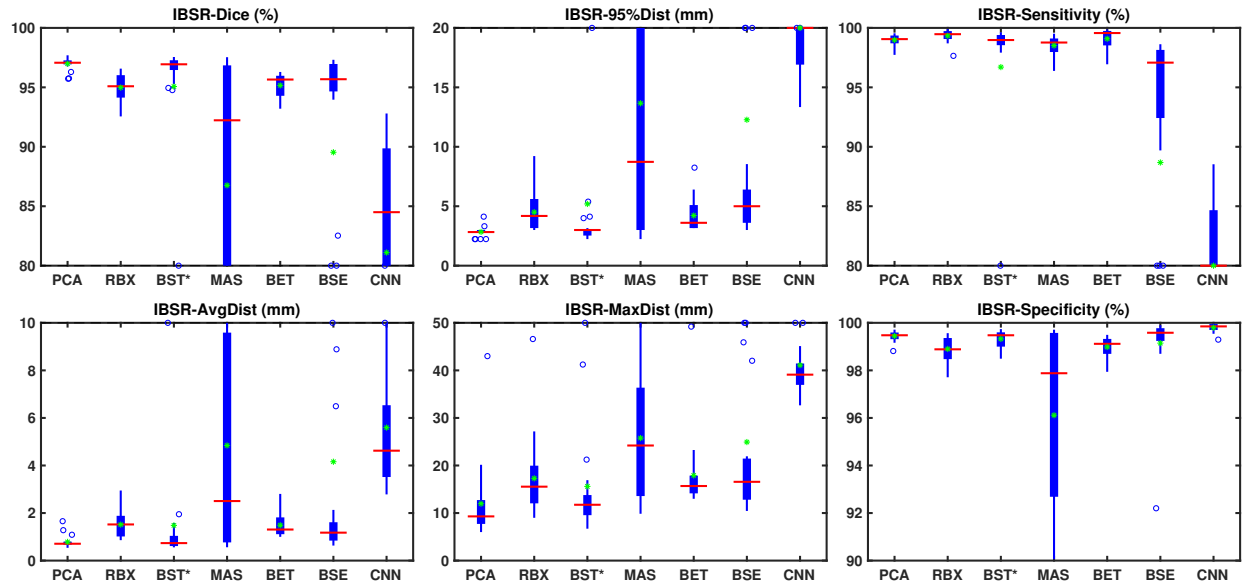


Figure 5.4: Box plot results for the IBSR dataset. Results from seven methods are shown: PCA, RBX (ROBEX), BST* (BEaST*), MAS (MASS), BET, BSE and CNN. Due to the poor results of MASS and CNN, and the outliers of BSE on this dataset, the range of the plots is limited for better visibility. On each box, the center line denotes the median, and the top and the bottom edge denote the 75th and 25th percentile, respectively. Outliers are marked with '+' signs. In addition, means are marked with green '*' signs. ROBEX, BET, and BSE show similar performance, but BSE exhibits two outliers. MASS works well on most images, but fails on many cases. BEaST fails on the original images. Thus the BEaST* results are shown using the initial affine registration from my PCA model. BEaST* performs well with high Dice scores and low surface distances, but with low mean values. CNN performs poorly on this dataset. The proposed PCA model has similar performance to BEaST* but with higher mean values. Both methods perform better than other methods on the Dice scores and surface distances.

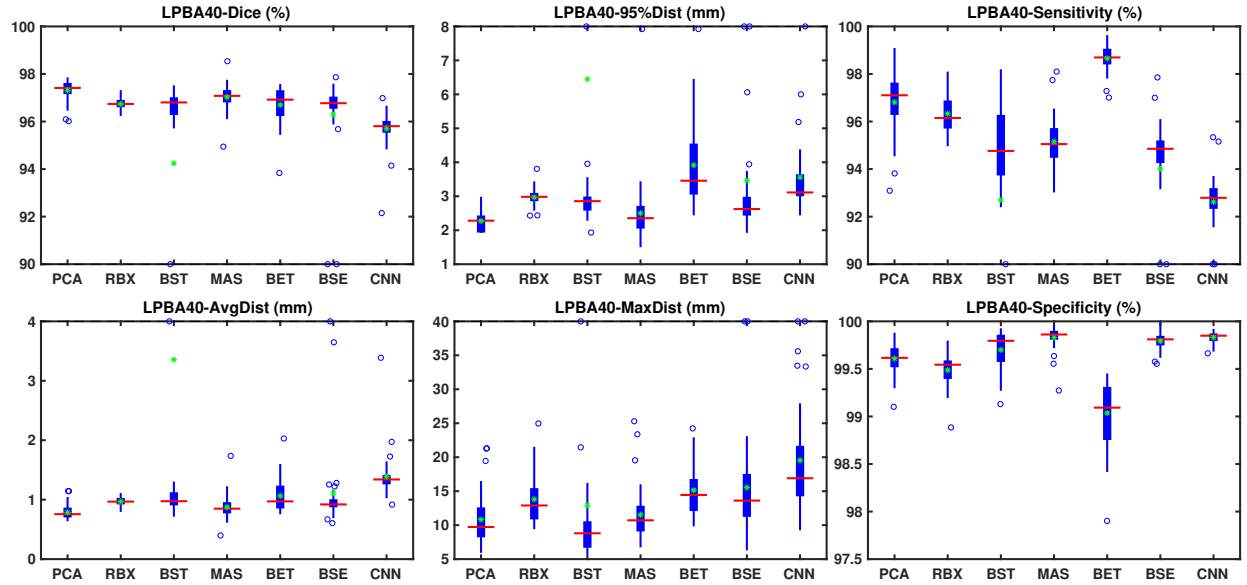


Figure 5.5: Box plot results for the LPBA40 dataset. All seven methods work well on this dataset. The proposed PCA model has the best Dice and surface distances. ROBEX, BEaST, MASS, BET and BSE show similar performance, but BET exhibits larger variance and BSE exhibits two outliers indicating failure. The CNN model shows overall slightly worse performance than the other methods.

in the pre-processing step for the proposed PCA model for all images. This affine transformation corresponds to a composition of the two affine transformations in Fig. 5.3. BEaST is then applied to the affinely aligned images. The same strategy is used for BraTS. The resulting approach is referred to as BEaST*. Overall, ROBEX, BEaST*, BSE, BET and the PCA model perform well on this dataset, with a median Dice coefficient above 0.95. BEaST* performs well on most cases with high Dice scores and low surface distances. MASS works well on some cases, but performs poorly on many cases. CNN does not perform satisfactorily, with low Dice scores, low sensitivity, large distance errors, and overall high variance. The proposed PCA model has similar performance to BEaST*, but does not result in extreme outliers and hence results in higher mean Dice scores than BEaST*. Both methods outperform all others with respect to Dice scores (median close to 0.97) and distance measures in most cases. BSE also works well on most cases, but it shows larger variability and exhibits two outliers which represent failure cases. ROBEX and BET show the highest sensitivity, but reduced specificity. Conversely, the proposed PCA model, BEaST*, BSE, and CNN have high specificity but reduced sensitivity (the CNN model dramatically so).

Table 5.1 (top) shows medians, means and standard deviations for the test results on this dataset. The proposed PCA model achieves the highest median and mean Dice overlap scores (both at 0.97) with the smallest standard deviation. BEaST* also shows high median Dice scores, but results in reduces mean scores due to the presence of outliers. ROBEX and BET show slightly reduced Dice overlap measures (mean and median around 0.95). BSE also shows slightly reduced median Dice scores, but greatly reduced mean scores. MASS show reduced median Dice scores. CNN shows the lowest performance. The proposed PCA model also performs best for the surface distance measures; it has the lowest mean and median surfaces distances. Overall the PCA model performs best.

In addition, one-tailed paired Wilcoxon signed-rank tests are performed (to safeguard against deviations from normality) to compare results between methods. The null hypothesis is that the paired differences for the results of the proposed model and of the compared method come from a distribution with zero median, against the alternative that the median of the paired differences

IBSR	PCA	ROBEX	BEaST*	MASS	BET	BSE	CNN
Dice(%)	97.07 96.99±0.53	95.09 94.98±1.17	96.94 95.07±7.50	92.23 86.76±11.06	95.66 95.16±0.96	95.68 89.54±21.76	84.50 81.10±12.07
Avg Dist(mm)	0.71 0.79±0.27	1.52 1.51±0.56	0.74 1.48±2.76	2.50 4.84±4.54	1.31 1.49±0.47	1.18 4.16±10.53	4.62 5.59±3.10
95% Dist(mm)	2.83 2.84±0.43	4.18 4.50±1.58	3.00 5.19±9.79	8.73 13.66±11.81	3.61 4.22±1.39	5.00 12.27±20.83	20.05 22.25±9.41
Max Dist(mm)	9.30 11.97±8.14	15.55 17.30±8.40	11.74 15.60±13.41	24.20 25.76±12.94	15.68 17.91±7.85	16.57 24.93±23.32	39.10 41.10±8.14
Sensitivity(%)	99.06 98.99±0.46	99.47 99.33±0.54	98.98 96.70±10.12	98.77 98.52±0.77	99.57 99.09±0.93	97.08 88.68±22.86	74.87 70.96±15.89
Specificity(%)	99.48 99.44±0.21	98.89 98.90±0.51	99.48 99.32±0.37	97.88 96.11±3.81	99.12 98.98±0.46	99.58 99.15±1.67	99.85 99.80±0.19
LPBA40	PCA	ROBEX	BEaST	MASS	BET	BSE	CNN
Dice(%)	97.41 97.32±0.42	96.74 96.74±0.24	96.80 94.25±15.29	97.08 97.03±0.57	96.92 96.70±0.78	96.77 96.29±2.26	95.80 95.70±0.74
Avg Dist(mm)	0.76 0.79±0.12	0.97 0.97±0.07	0.97 3.36±14.83	0.85 0.88±0.21	0.97 1.06±0.27	0.92 1.11±0.81	1.34 1.39±0.37
95% Dist(mm)	2.28 2.27±0.32	2.98 2.97±0.26	2.86 6.45±23.19	2.36 2.50±0.97	3.46 3.92±1.24	2.62 3.46±3.38	3.11 3.56±1.56
Max Dist(mm)	9.73 10.83±3.76	12.89 13.81±3.47	8.80 12.89±24.44	10.71 11.53±4.04	14.44 15.14±3.75	13.61 15.54±7.74	16.91 19.55±8.17
Sensitivity(%)	97.10 96.81±1.23	96.15 96.33±0.85	94.76 92.70±15.12	95.05 95.15±1.08	98.70 98.66±0.54	94.85 94.02±4.10	92.79 92.62±1.46
Specificity(%)	99.62 99.61±0.16	99.54 99.49±0.16	99.80 99.70±0.21	99.86 99.83±0.12	99.09 99.04±0.34	99.81 99.79±0.09	99.85 99.83±0.07
BraTS	PCA	ROBEX	BEaST*	MASS	BET	BSE	CNN
Dice(%)	96.34 96.16±0.92	95.15 94.83±1.49	94.99 93.29±7.00	96.20 95.71±1.39	94.40 90.95±13.41	84.21 84.91±8.89	1.75 21.89±29.54
Avg Dist(mm)	0.97 1.00±0.31	1.31 1.54±0.70	1.38 2.28±3.34	1.03 1.17±0.56	1.68 7.58±25.30	4.65 4.37±3.61	55.88 44.87±29.05
95% Dist(mm)	4.15 4.35±1.27	5.23 6.03±2.50	5.01 7.71±9.29	4.62 4.87±2.04	5.87 6.18±3.53	13.52 13.92±13.00	78.45 73.85±38.77
Max Dist(mm)	12.03 15.26±9.32	12.83 16.42±8.80	15.03 20.32±14.57	13.29 16.43±8.97	17.49 22.78±22.61	33.95 32.02±22.38	87.73 86.60±36.92
Sensitivity(%)	96.16 96.17±1.84	95.82 94.95±2.88	97.45 97.06±2.66	94.46 93.62±2.87	94.92 94.77±3.82	74.28 77.80±13.43	0.89 16.17±24.73
Specificity(%)	99.33 99.29±0.25	99.16 98.98±0.65	98.68 97.28±5.46	99.73 99.62±0.28	98.90 93.69±22.08	99.82 99.29±2.38	100.00 99.97±0.05
TBI	PCA	ROBEX	BEaST	MASS	BET	BSE	CNN
Dice(%)	96.40 96.28±0.85	93.71 93.60±1.00	95.04 95.06±0.96	95.11 95.42±0.96	95.40 95.14±1.12	92.64 91.00±4.31	91.51 90.40±5.07
Avg Dist(mm)	1.20 1.22±0.30	2.12 2.20±0.40	1.60 1.62±0.33	1.61 1.53±0.33	1.50 1.57±0.40	2.23 3.15±1.66	3.20 3.46±1.75
95% Dist(mm)	3.37 3.41±0.85	6.20 5.99±0.97	6.24 5.86±1.44	4.90 4.59±0.77	5.06 5.57±1.91	9.46 13.07±7.11	14.97 16.04±8.72
Max Dist(mm)	16.13 15.54±5.03	18.25 18.89±5.12	15.09 17.27±4.60	15.65 16.08±4.16	17.46 18.53±4.59	27.16 31.96±12.71	41.49 37.06±10.09
Sensitivity(%)	97.82 97.76±0.92	98.91 98.64±0.93	92.12 92.89±2.44	98.74 98.68±0.66	93.98 93.65±1.87	87.51 85.65±8.17	85.69 83.77±8.58
Specificity(%)	99.03 99.07±0.26	98.09 97.75±0.93	99.63 99.57±0.22	98.64 98.59±0.47	99.41 99.44±0.11	99.82 99.54±0.86	99.80 99.81±0.07

Table 5.1: Medians (top), and means with standard deviations (bottom) for validation measures for all the methods and all the datasets. The best results are highlighted in green based on the *median* values. Among all datasets, the proposed PCA model has the best median on Dice overlap scores and generally on surface distances. Exception is BEaST which achieves a lower maximum surface distances on the LPBA40 and the TBI datasets. In addition, the proposed model also has the best mean and variance for the Dice overlap scores and the surface distances on most of these datasets.

IBSR	ROBEX	BEaST*	MASS	BET	BSE	CNN
Dice	4.78e-5	1.20e-2	2.77e-4	4.78e-5	7.55e-5	4.78e-5
Avg Dist	4.78e-5	2.73e-2	1.82e-4	4.78e-5	4.78e-5	4.78e-5
95% Dist	4.74e-5	5.91e-2	1.05e-4	4.71e-5	4.74e-5	4.78e-5
Max Dist	4.78e-5	5.36e-2	4.78e-5	4.78e-5	1.58e-4	5.58e-5
Sensitivity	0.994	0.448	3.40e-3	0.829	4.78e-5	4.78e-5
Specificity	5.58e-5	2.97e-2	2.41e-3	4.78e-5	0.894	1.000

LPBA40	ROBEX	BEaST	MASS	BET	BSE	CNN
Dice	1.47e-7	2.51e-8	1.89e-3	9.58e-5	2.24e-7	1.85e-8
Avg Dist	1.36e-7	2.51e-8	2.75e-3	1.60e-6	6.31e-7	1.85e-8
95% Dist	2.90e-8	3.29e-7	5.69e-2	2.71e-8	1.02e-5	1.85e-8
Max Dist	2.16e-8	1.000	2.58e-2	2.92e-8	3.01e-5	2.51e-8
Sensitivity	4.13e-3	1.27e-7	1.60e-6	1.000	6.14e-8	1.85e-8
Specificity	5.70e-6	0.998	1.000	2.00e-8	1.000	1.000

BraTS	ROBEX	BEaST*	MASS	BET	BSE	CNN
Dice	1.58e-4	3.18e-4	7.02e-2	4.78e-5	4.78e-5	4.78e-5
Avg Dist	1.36e-4	2.77e-4	9.89e-2	4.78e-5	4.78e-5	4.78e-5
95% Dist	8.41e-5	4.17e-4	0.266	1.53e-3	4.78e-5	7.15e-5
Max Dist	1.91e-2	7.38e-4	0.222	2.41e-4	1.18e-4	4.78e-5
Sensitivity	3.51e-2	0.981	2.09e-4	8.08e-2	5.58e-5	4.78e-5
Specificity	6.53e-2	1.82e-4	0.999	4.73e-3	0.999	1.000

TBI	ROBEX	BEaST	MASS	BET	BSE	CNN
Dice	3.91e-3	1.95e-2	2.73e-2	7.81e-3	3.91e-3	3.91e-3
Avg Dist	3.91e-3	1.95e-2	3.91e-2	7.81e-3	3.91e-3	3.91e-3
95% Dist	3.91e-3	7.81e-3	7.81e-3	3.91e-3	3.91e-3	3.91e-3
Max Dist	1.17e-2	9.77e-2	0.344	5.47e-2	3.91e-3	3.91e-3
Sensitivity	0.980	3.91e-3	0.961	3.91e-3	3.91e-3	3.91e-3
Specificity	3.91e-3	1.000	2.73e-2	1.000	0.926	1.000

Table 5.2: p-values for all datasets, computed by signed-rank tests. One-tailed paired Wilcoxon signed-rank tests are performed, where the null-hypothesis ($\mathcal{H}_0$) is that the paired differences for the results of the proposed PCA model and of the compared method come from a distribution with zero median, against the alternative ($\mathcal{H}_1$) that the paired differences have a non-zero median (greater than zero for Dice, sensitivity and specificity, and less than zero for surface distances). In addition, the Benjamini-Hochberg procedure is used to reduce the false discovery rate (FDR). The results are highlighted in green, where the proposed PCA model performs statistically significantly better. The results show that proposed PCA model outperforms other methods on most of the measures.

is non-zero.[5]. Table 5.2 (top) shows the corresponding results. The Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) is applied for all the tests, in order to reduce the false discovery rate for multiple comparisons. An overall false discovery rate of 0.05 is selected which results in an effective significance level of $\alpha \approx 0.0351$. The proposed model outperforms all other methods on Dice and surface distances except for BEaST* which is significant only in Dice and average surface distance. In addition, it performs better than MASS, BSE and CNN on sensitivity and better than ROBEX, BEaST*, MASS, and BET on specificity.

LPBA40 results: Figure 5.5 shows the box-plots summarizing the validation results for the LPBA40 dataset. All seven methods perform well. ROBEX, BEaST, BET and BSE all have a median Dice score between 0.96 and 0.97. MASS has a median Dice score slightly above 0.97. The proposed PCA model obtains the highest median Dice score (0.974). All methods except for the CNN approach have a median average surface distance smaller than 1 mm. Table 5.1 (second top) shows the medians, means and standard deviations for all validation measures for this dataset. Again, all methods have satisfactory median, mean Dice scores and surface distances with low variances. Compared with other methods, the PCA model achieves the best results.

Table 5.2 (second top) shows the one-sided paired Wilcoxon signed-rank test results. Again the Benjamini-Hochberg procedure is used. All methods perform well on this dataset, but the proposed PCA approach still shows statistically significant improvement. Specifically, it outperforms other methods on Dice and all surface distances with statistical significance except for BEaST on maximum surface distance and for MASS on 95% surface distance. It also performs better than all other methods except BET on sensitivity and better than BET and ROBEX on specificity.

Figure 5.6 visualizes the average brain mask errors for IBSR and LPBA40. All images are first affinely registered to the atlas. The gold-standard expert segmentations are transformed as well as the automatically obtained brain masks of the different methods to atlas space. The segmentations are compared by counting the average over- and under-segmentation errors over all cases at each voxel. This results in a visualization for areas of likely mis-segmentation. The proposed model,

[5]One-tailed tests are performed, thus they are for greater than zero for the Dice overlap scores, sensitivity and specificity, and less than zero for the surface distances.

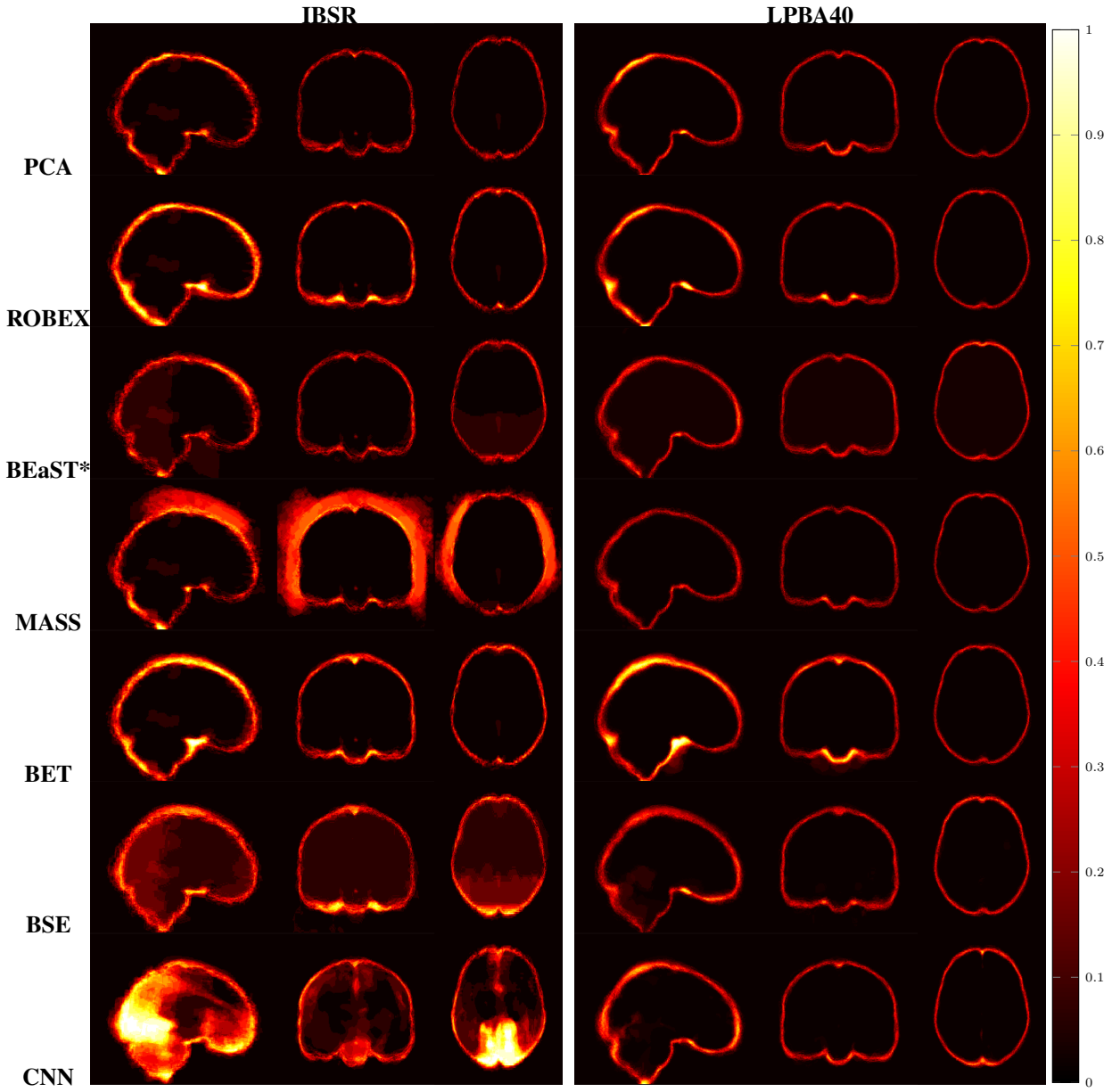


Figure 5.6: Examples of 3D volumes of average errors for the normal IBSR and LPBA40 datasets. For IBSR, results for BEaST* are shown. Images and their brain masks are first affinely aligned to the atlas. At each location the proportion of segmentation errors are then calculated among all the segmented cases of a dataset (both over- and under-segmentation errors). Lower values are better (a value of 0 indicates perfect results over all images) and higher values indicate poorer performance (a value of 1 indicates failure on all cases). BSE and CNN struggle with the IBSR dataset. MASS also shows large errors near the boundary. The PCA method shows good performance on normal datasets.

ROBEX, BEaST (BEaST*) and BET perform well on these two datasets. Compared to the PCA model, ROBEX, BEaST (BEaST*) and BET show larger localized errors, e.g., at the boundary of the parietal lobe, the occipital lobe and the cerebellum. While MASS, BSE and CNN perform well on the LPBA40 dataset, they perform poorly on the IBSR dataset. This is in particular the case for the CNN approach.

5.3.4 Datasets with strong pathologies: BraTS/TBI

BraTS results: Figure 5.7 shows the box-plots for the validation measures for the BraTS dataset. BSE and CNN, using their default settings, do not work well on the BraTS dataset. This may be because of the data quality of the BraTS data. Many of the BraTS images have relatively low out-of-plane resolutions. BSE results may be improved by a better parameter setting. However, as my goal is to evaluate all methods with the same parameter setting across all datasets, I do not explore dataset specific parameter tuning. BEaST also fails on the original BraTS images due to the spatial normalization. As for the IBSR dataset, BEaST* is therefore used, which is the adaptation of BEaST using the affine transformation of the PCA model. BET shows good performance, but suffers from a few outliers. ROBEX and BEaST* work generally well, with a median Dice score around 0.95 and an average distance error of 1.3 mm. MASS also works well on most cases. However, as for IBSR and LPBA40, the proposed PCA model performs generally the best with a median Dice score 0.96 and a 1 mm average distance error. It also shows lower variance, as shown in table 5.1 (second bottom), underlining the very consistent behavior of the proposed approach.

Table 5.2 shows (via a one-sided paired Wilcoxon signed-rank test with a correction for multiple comparisons using a false discovery rate of 0.05) that the proposed model has statistically significantly better performance than ROBEX, BEaST*, BET, BSE, CNN on most measures. The improvement over MASS, however, is not statistically significant.

TBI results: Figure 5.8 shows the box-plots for the results on the TBI dataset. The proposed PCA model still outperforms all other methods. It achieves the largest Dice scores, and the lowest surface distances among all methods with best mean and lowest variance as shown in Table 5.1 (bottom). Table 5.2 shows the one-sided paired Wilcoxon signed-rank test results with multiple comparisons

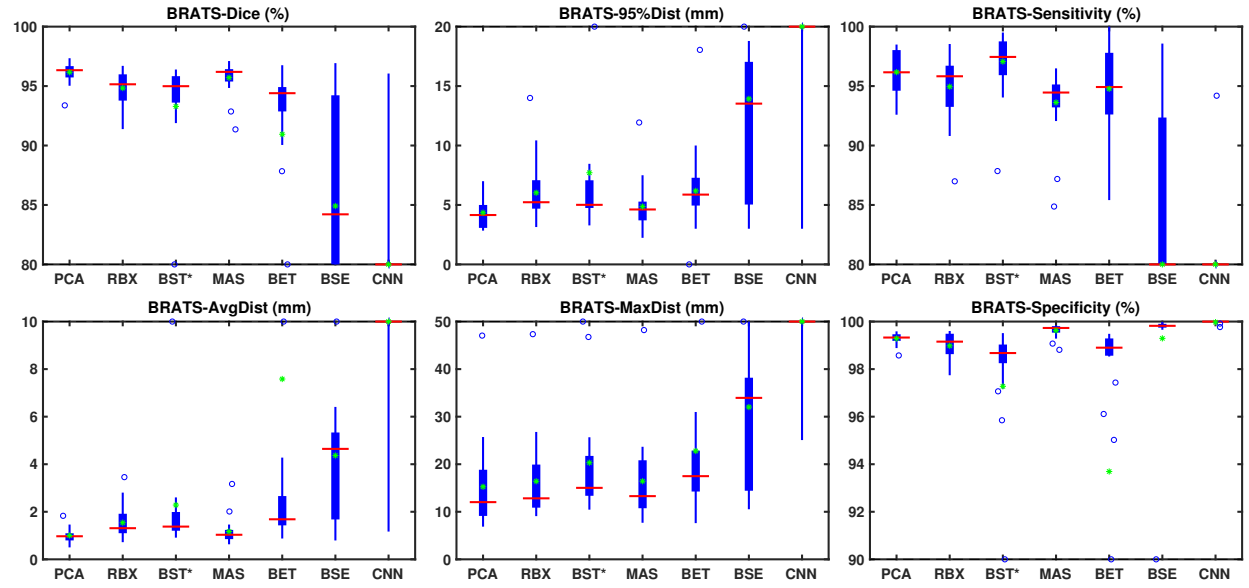


Figure 5.7: Box plot results for the BraTS tumor dataset. BSE and CNN fail on this dataset. BEaST also fails when applied directly to the BraTS dataset due to spatial normalization failures. Thus results for BEaST* are shown here, which is the modification using the affine registration of the PCA model first. BET shows better performance, but also exhibits outliers. ROBEX, BEaST*, MASS, and the proposed PCA model work well on this dataset. Overall the PCA model exhibits the best performance scores.

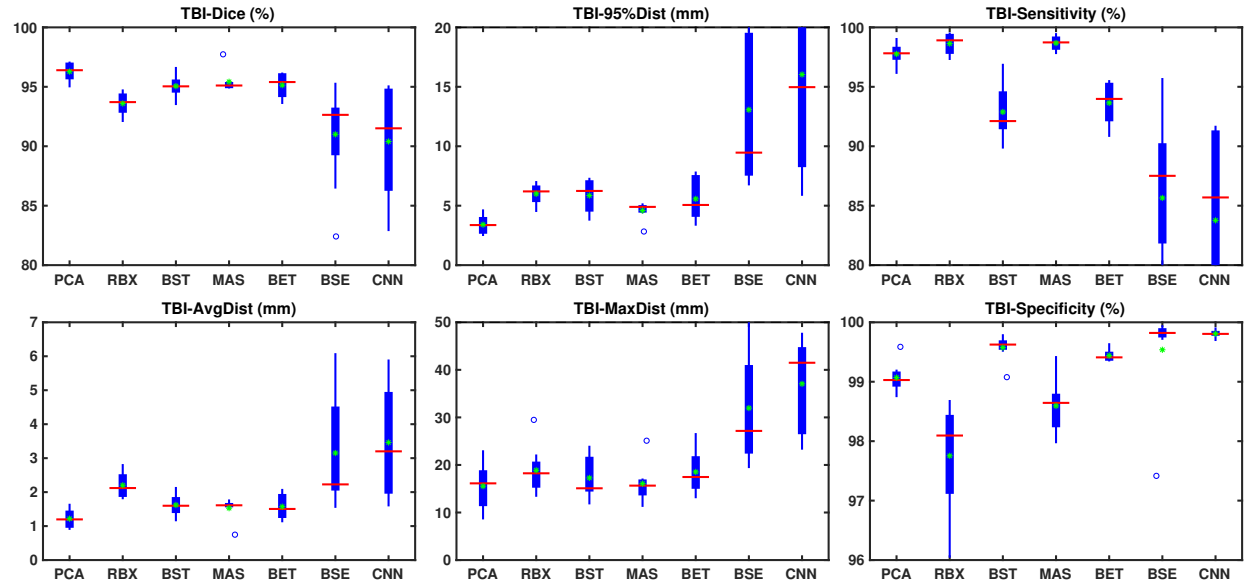


Figure 5.8: Box plot results for the TBI dataset. The proposed PCA model shows the best evaluation scores. BET, BEaST, MASS and ROBEX also perform reasonably well. BSE and CNN exhibit inferior performance on this dataset.

correction with a false discovery rate of 0.05. The proposed model performs significantly better than ROBEX, BEaST, BET, BSE and CNN on most measures. The improvement over MASS is only statistically significant on Dice and 95% surface distance.

Finally, Figure 5.9 shows the average segmentation errors on the BraTS and TBI datasets: the PCA method shows fewer errors than most other methods in these two abnormal datasets. MASS also shows few errors, while ROBEX, BEaST (BEaST*) and BET exhibit slightly larger errors at the boundary of the brain. CNN and BSE particularly show large errors for the BraTS dataset presumably again due to the coarse resolution of the BraTS data.

In addition to extracting the brain from pathological datasets, the proposed method also allows for the estimation of a corresponding quasi-normal image in atlas space, although this is not the main goal of this work. Figure 5.10 shows an example of the reconstructed quasi-normal image (L) for an image of the BraTS dataset, as well as an estimation of the pathology (pathology image T and non-brain image S). Compared to the original image, the pathology shown in the quasi-normal image has been greatly reduced. Hence this image can be used for the registration with a normal image or a normal atlas. This has been shown to improve registration accuracy for the registration of pathological images in Chapter 3. Furthermore, an estimate of the pathology (here a tumor) is also obtained which may be useful for further analysis. Note that in this example image the total variation term captures more than just the tumor. This may be due to inconsistencies in the image appearance between the normal images (obtained from OASIS data) and the test dataset. As my goal is atlas alignment rather than quasi-normal image reconstruction or pathology segmentation, such a decomposition is acceptable, although it could be improved by tuning the parameters or applying regularization steps as in Chapter 3.

5.3.5 Runtime and memory consumption

Decomposition is implemented on the GPU. Each decomposition takes between 3 to 5 minutes. Currently, the registration steps are the most time-consuming parts of the overall algorithm. `NiftyReg` is used on the CPU for registrations. Each affine registration step takes less than 3 minutes and the B-spline step takes 5 minutes. However, in the current version of `NiftyReg` a

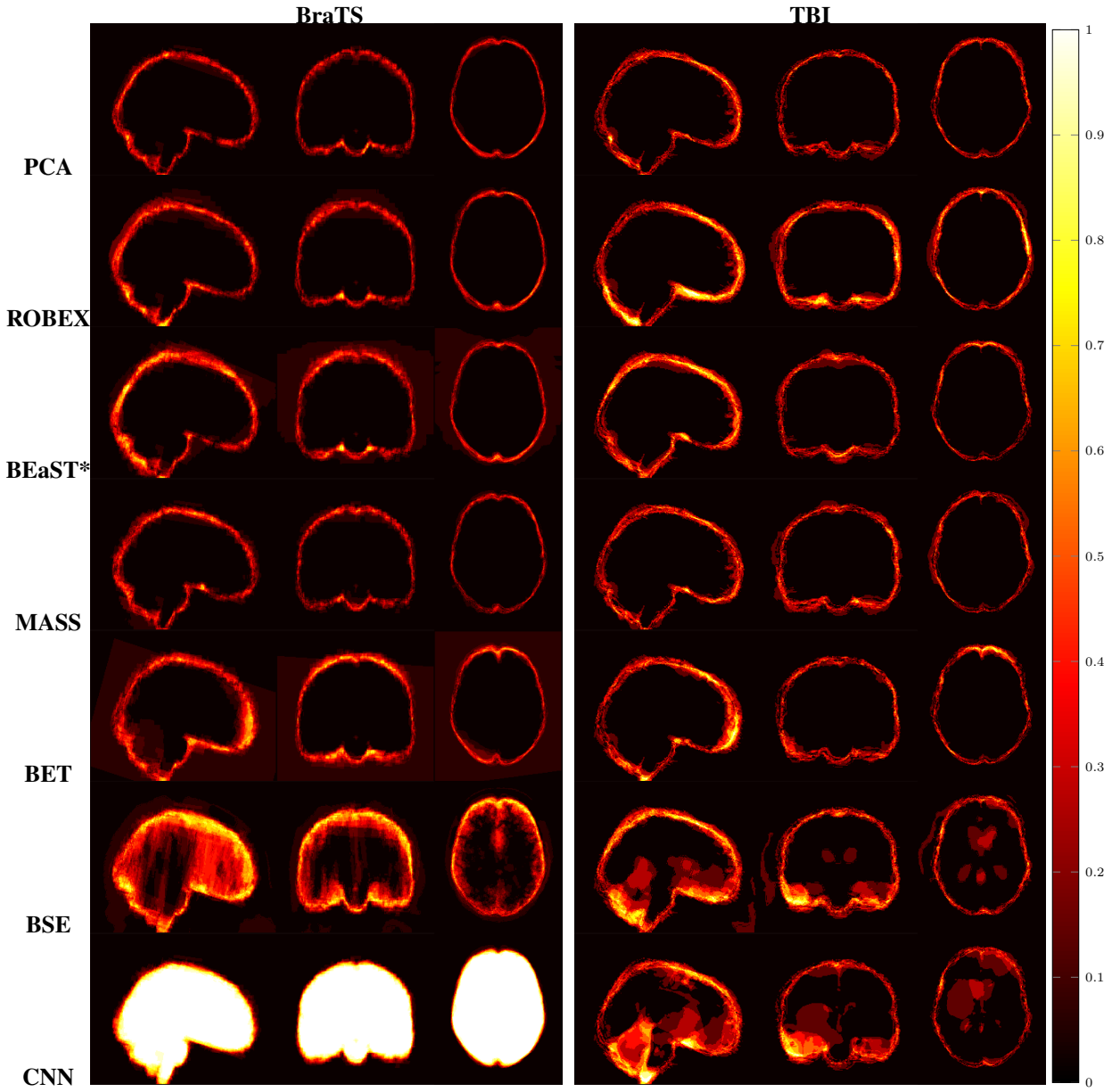
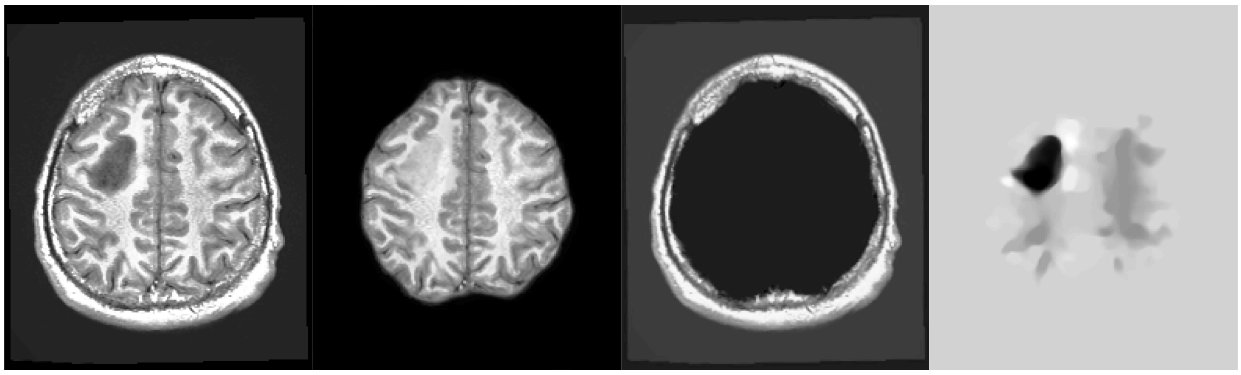


Figure 5.9: Examples of 3D volumes of average errors for the pathological BraTS and TBI datasets. For BraTS, results for BEaST* are shown. Images and their brain masks are first affinely aligned to the atlas. At each location the proportion of segmentation errors are then calculated among all the segmented cases of a dataset (both over- and under-segmentation errors). Lower values are better (a value of 0 indicates perfect results over all images) and higher values indicate poorer performance (a value of 1 indicates failure on all cases). BSE and CNN fail on the BraTS and TBI dataset whereas the PCA method shows excellent performance.



(a) (b) (c) (d)

Figure 5.10: Example BraTS image with its decomposition result in atlas space. (a) Input image after pre-processing; (b) quasi-normal image $L + M$; (c) non-brain image S; (d) pathology image T.

B-spline registration can take up to 15 minutes when cost function masking is used. Overall the proposed brain extraction approach takes around 1 hour to 1.5 hours for each case, including the pre-processing step.

Storing the PCA basis requires the most memory. Each 197×232×189 3D image (stored as double) consumes about 66MB of memory. Hence it requires less than 7 GB to store the 100 PCA basis images, in addition to the atlases and masks. As the PCA model only uses 50 PCA bases, stored in B, and requires two variable copies during runtime, the overall algorithm requires less than 7 GB of memory and hence can easily be run on modern GPUs.

The software is freely available as open source code at `https://github.com/uncbiag/pstrip`.

5.4 Discussion

In this chapter, I presented a PCA-based model specifically designed for brain extraction from pathological images. The model decomposes an image into three parts. Non-brain tissue outside of the brain is captured by a sparse term, normal brain tissue is reconstructed as a quasi-normal image close to a normal PCA space, and brain pathologies are captured by a total-variation term. The quasi-normal image allows for registration to an atlas space, which in turn allows registering the original image to atlas space and hence to perform brain extraction.

While the proposed approach is designed for reliable brain extraction from images with strong pathologies, it also performs well for normal images or images with subtle pathologies. This is in contrast to most of the existing methods, which assume normal images or only slight pathologies. These algorithms are either not designed for pathological data (BET, BSE, BEaST) or use normal data for training (e.g., ROBEX and CNN). Consequently, as demonstrated in the experiments, these methods may work suboptimally or occasionally fail when presented with pathological data. While the proposed PCA model is built on OASIS data, which contains abnormal images (from patients with Alzheimer's disease), OASIS data does not exhibit strong pathologies as, for example, seen in the BraTS and the TBI datasets. However, as the algorithm is specifically modeling pathologies on top of a statistical model of *normal* tissue appearance, it can tolerate pathological data better and, in particular, does not require pathology-specific training.

One of the main advantages of the proposed brain extraction method is that a *fixed* set of parameters (without additional tuning or dataset-specific brain templates) can be used across a wide variety of datasets. This can, for example, be beneficial for small-scale studies, where obtaining dataset-specific templates may not be warranted, or for more clinically oriented studies, where image appearance may be less controlled.

On the other hand, this generality likely implies suboptimality. For example, a likely reason why the CNN approach performs poorly on some of the datasets is that these datasets do not correspond well to the data the CNN was trained on. Dataset-specific fine-tuning of the model would likely help improve the CNN performance. Similarly, approaches, including the proposed, relying on some form of registration, and a model would likely benefit from a dataset-specific atlas (including a dataset-specific PCA basis in my case) or dataset-specific registration templates. Such dataset-specific templates can, for example, easily be used within MASS and improve performance slightly. Similarly, I observed that the performance for BEaST could be improved if dataset-specific libraries are used. In practice, large-scale studies may warrant the additional effort of obtaining dataset-specific manually segmented brain masks for training. However, in many cases, such manual segmentation may be too labor-intensive. In this latter case, the proposed approach is particularly

attractive as it is only moderately affected by differing image appearances and works well with a generic model for brain extraction.

Tumors or general pathologies may also affect some of the pre-processing steps. For example, histogram matching is performed over the entire initial brain mask, which includes the pathology. In practice, it has been visually assessed that such a histogram matching strategy produced reasonable intensity normalizations. However, this step could be improved, for example, by coupling it or alternating it with the decomposition in such a way that regions that likely correspond to pathologies are excluded from the histogram computations for histogram matching.

CHAPTER 6: A Deep Network for Joint Registration and Reconstruction of Images with Pathologies

In this chapter, I present a joint registration and reconstruction network for an image with pathology. The presented network learns to reconstruct a quasi-normal image from an image with tumors, while simultaneously predicting the transformation to an atlas space. Using separate decoders, tumor mass effects and the reconstruction of quasi-normal images are learned separately. Contributions for this chapter are as follows:

- *Joint reconstruction and registration network.* To the best of my knowledge, this is the first deep network trained jointly to reconstruct and register a brain image with an included pathology to an atlas. The network recovers the missing correspondences between the pathologies and the atlas space. It is also more computationally efficient than previous approaches by avoiding the interleaving of registrations and reconstructions, resulting in rapid predictions at test time.

- *Reconstruction of quasi-normal appearance in atlas space.* As the transformation to the atlas is disentangled from the reconstruction, quasi-normal image appearance is learned in atlas space, which simplifies the appearance modeling.

- *Vector-momentum parameterized fluid-based registration.* The network incorporates a vector-momentum parameterized stationary velocity field (vSVF) (Shen et al., 2019a), which can capture large deformations while retaining diffeomorphic transformations. Instead of the input image, the reconstructed quasi-normal image is used to drive the registration, which provides better and more reliable guidance.

- *Validation.* Results on both synthetic and real brain tumor scans show that the proposed network successfully learns to reconstruct quasi-normal appearance simultaneously with the transformation of the tumor image to atlas space. Specifically, improvements are shown over cost function masking, which demonstrates that modeling quasi-normal image structure is beneficial for the registration of images with pathologies.

The remainder of this chapter is organized as follows: Section 6.1 describes the proposed registration and reconstruction network, including its architecture and associated loss functions, and section 6.2 presents experimental details and results on both a synthetic brain tumor dataset, and on a paired set of pre-operative and post-recurrence brain tumor scans. Section 6.3 concludes this chapter with a summary. The work presented in this chapter has been accepted for publish in 2020 Machine Learning in Medical Imaging (MLMI 2020) (Han et al., 2020).

6.1 Joint Registration and Reconstruction Network

Fig. 6.1 shows an overview of the proposed network. The network takes a tumor image I_T and an atlas A as its inputs and outputs a vector-momentum parameterization of the transformation Φ^{-1}, a reconstructed quasi-normal image I_R and a segmentation of the tumor region I_S. The network jointly learns both the registration and reconstruction, which is more efficient than approaches that interleave registrations and reconstructions. Importantly, the transformation warps the tumor image to the atlas for a better reconstruction in *atlas space*, while the *reconstructed image* guides the similarity measure so that the network learns a better transformation as it is no longer perturbed by the pathology.

6.1.1 Registration Decoder

A vector-momentum parameterized stationary velocity field (vSVF) is used in the proposed model (Shen et al., 2019a; Niethammer et al., 2019). Instead of directly predicting the transformation field, the network predicts a momentum vector field, m, which gets smoothed by a multi-Gaussian kernel (Risser et al., 2011) resulting in a velocity vector field, v, from which the transformation map,

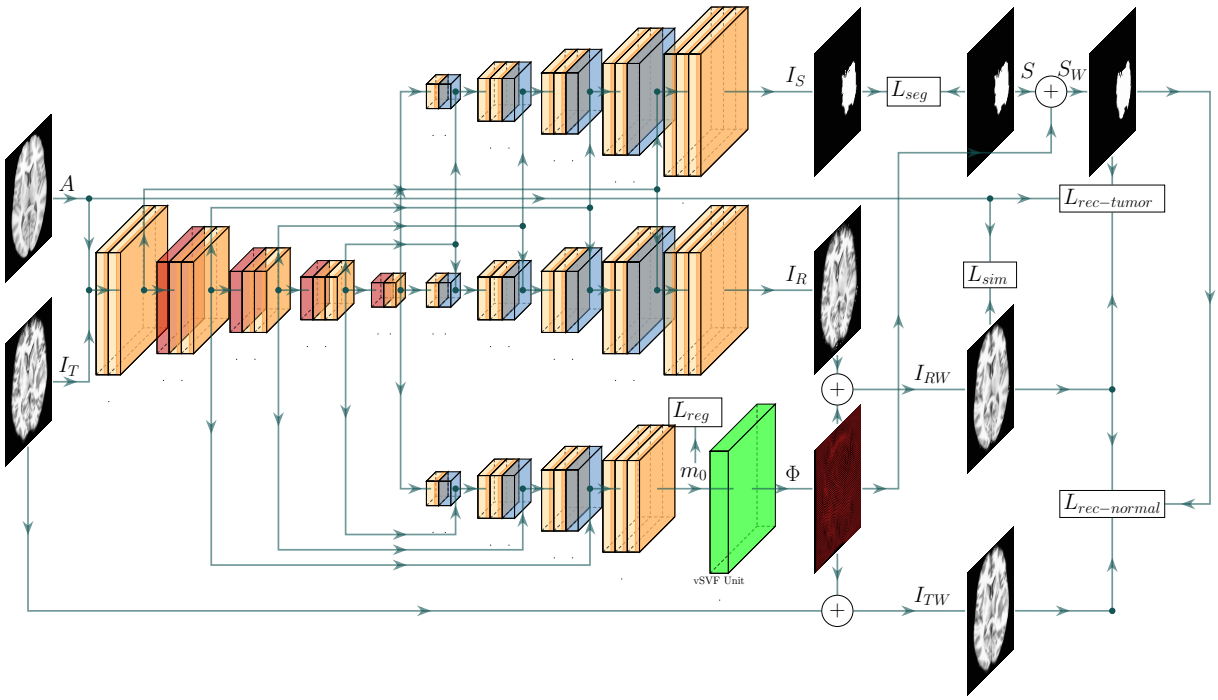


Figure 6.1: Overview of the proposed network. The atlas A, the tumor image I_T and the tumor mask S are given. The network outputs a mask I_S, a reconstructed quasi-normal image I_R and a vector momentum m_0 which is used to obtain the transformation map Φ^{-1}. A vector momentum-parametrized stationary velocity field (vSVF) is used as the registration model. The regularization loss L_{reg} penalizes m_0, while the similarity loss L_{sim} penalizes the warped reconstructed image I_{RW} with respect to the atlas A. They are backpropagated through the registration decoder. The reconstruction loss penalizes the warped quasi-normal image in the tumor region and the normal region, given by $L_{rec-normal}$ and $L_{rec-tumor}$. $L_{rec-normal}$ penalizes the image difference between the warped tumor image I_{TW} and the warped quasi-normal image I_{RW} in the normal region and $L_{rec-tumor}$ penalizes the image difference between the warped quasi-normal image and the atlas in the tumor region. Backpropagation is through the reconstruction decoder. The segmentation loss L_{seg} takes the predicted mask and backpropagates through the segmentation decoder.

Φ^{-1}, is computed via integration. The benefit of this indirect way is that it can assure diffeomorphic transformations at *test* time.

The registration loss consists of a regularization loss and a similarity loss:

$$\begin{aligned} L_{rgs}(m_0) =& \langle m_0, v_0 \rangle + \frac{1}{\sigma^2} Sim[I_R \circ \Phi^{-1}(1), A], \\ \text{s.t.} \quad & \Phi_t^{-1} + D\Phi^{-1} v_0 = 0, \quad \Phi^{-1}(0) = \Phi_{(0)}^{-1}, \quad v_0 = (L^\dagger L)^{-1} m_0, \end{aligned} \tag{6.1}$$

where D denotes the Jacobian, m_0 is the initial vector momentum, $\sigma > 0$ balances the two terms, and $\Phi_{(0)}^{-1}$ is the initial condition for the transformation map, Φ^{-1}, which can be set to identity or to the

transformation of a pre-registration, for example, an affine registration. Localized normalized cross correlation (LNCC) is used as the similarity loss as in (Shen et al., 2019a). A significant difference from existing registration networks is that instead of using the input tumor image I_T to evaluate the similarity loss, the reconstructed image I_R is used, which is output via the reconstruction decoder. The reconstructed I_R recovers image correspondences which can guide image registration. The registration loss only backpropagates through the registration decoder.

6.1.2 Reconstruction Decoder

The reconstruction decoder predicts a quasi-normal image from the tumor image. This mapping is directly learned from the atlas appearance. Specifically, for a given tumor image, its manually segmented tumor mask, S, is used to separate the tumor and the normal region (the tumor mask is only used during training.). In the normal region, the warped reconstruction image $I_{RW} = I_R \circ \Phi^{-1}(1)$ should be close to the warped original image $I_{TW} = I_T \circ \Phi^{-1}(1)$. In the tumor region, the reconstruction should be close to the atlas A. The warped tumor mask is $S_W = S \circ \Phi^{-1}(1)$. The reconstruction loss is defined as follows:

$$L_{rec} = \frac{1}{|\Omega_N|} \int_{\Omega_N} (I_{RW} - I_{TW})^2 \, dx + \frac{1}{|\Omega_T|} \int_{\Omega_T} (I_{RW} - A)^2 \, dx, \tag{6.2}$$

where $\Omega_N = \{x : S_W(x) = 0\}$ is the normal domain, $\Omega_T = \{x : S_W(x) = 1\}$ is the tumor domain, and $|\Omega|$ denotes the volume of domain Ω. The loss captures the sum of the mean-squared errors over the normal region and the tumor region. Atlas appearance is used to learn the tumor-to-quasi-normal mapping since the atlas is the target image. This can be considered a highly simplified statistical model only represented by its mean, the atlas. Combinations with more advanced statistical models, for example based on principal component analysis (Liu et al., 2014; Han et al., 2017) or variational autoencoders (Kingma and Welling, 2013), are conceivable. Moreover, as the primary goal of the network is for registration, simplifying the normal tissue modeling is a reasonable approach as long as the quasi-normal image can help to establish correspondences to the atlas. The reconstruction loss only backpropagates through the reconstruction decoder.

6.1.3 Segmentation Decoder

In principle, the segmentation decoder is not required for registration and reconstruction. Since the segmentation mask is used during training for reconstruction, a segmentation decoder is also added which outputs a predicted segmentation of the tumor. This is similar to (Shu et al., 2018), where an instance class can also be predicted. Intuitively, by providing direct supervision on the segmentation, the network is required to learn a representation capable of separating the tumor from the normal region. Binary cross-entropy loss is used, where the output of the segmentation decoder, I_S, is the predicted probability that a voxel belongs to the tumor region:

$$L_{seg} = -\frac{1}{|\Omega|}\int_{\Omega} S \log I_S + (1-S)\log(1-I_S)\, dx. \tag{6.3}$$

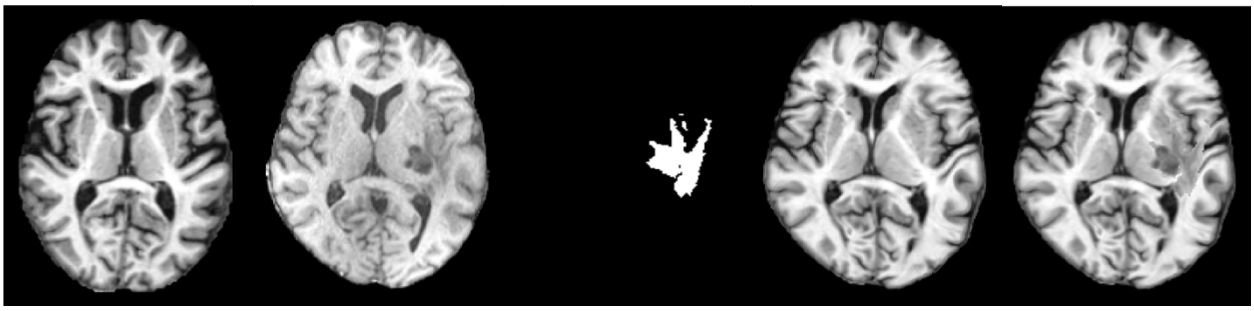
6.2 Experiments and Results

A pseudo-tumor dataset was created with a synthetic ground-truth for the reconstructions. In this section, I show that it is beneficial to use the quasi-normal image reconstructions for registration. Also, a dataset of pre-operative and post-recurrence magnetic resonance images (MRIs) is used from patients with glioblastomas with expert-placed landmarks for validation. I show that the predicted registration by the proposed network is more accurate than cost function masking and direct registration of the tumor images. ICBM 152 (Fonov et al., 2009) is used as the atlas.

6.2.1 3D Pseudo-tumor Dataset

The pseudo-tumor dataset is created using BraTS2019 (Menze et al., 2014; Bakas et al., 2017, 2018) and OASIS-3 (LaMontagne et al., 2019). OASIS-3 contains longitudinal MRIs from over 1,000 participants with normal cognitive function and with various stages of cognitive decline. The BraTS data contains MRIs from patients with brain tumors and corresponding tumor segmentations. Two hundred and eighty pairs of T1w-images are randomly selected; one from OASIS (only one scan for each patient is used) and one from BraTS. To mimic the mass effect of a brain tumor, the OASIS T1w scan was registered to the BraTS T1w scan with cost function masking. These warped

OASIS images are used as ground-truth images as they do not contain tumors but include mass effects. Brain tumors are then pasted from the BraTS scans onto the ground-truth scans (deformed OASIS scans), which results in pseudo-tumor images. The resulting 280 simulated images are the pseudo-tumor dataset. Forty are randomly selected 40 for testing, 40 for validation and 200 for training. Images are affinely aligned to the atlas, which is resampled to $128 \times 128 \times 128$ with $1.5 \times 1.5 \times 1.5$ mm^3 isotropic voxels. Figure 6.2 shows an example of the pseudo tumor.



OASIS Image | BraTS Image | BraTS Seg | Image w/o Tumor (Ground-truth) | Image w/ Tumor (Pseudo-tumor)

Figure 6.2: Example of a pseudo-tumor image. The OASIS image and the BraTS image are both intensity normalized to [0,1]. The BraTS segmentation is also provided in the dataset. The OASIS image is first registered to the BraTS image with cost function masking. The registration is to simulate the mass effect. The warped OASIS image is the ground-truth image. It does not contain the tumor, but includes the simulated mass effect. The BraTS tumor is then copied from the BraTS image and pasted onto the ground-truth image, which results in the pseudo-tumor image.

Since this dataset is simulated, ground-truth images are available which do not have tumors but include mass effects. The atlas is registered to these images. As these registrations are not impacted by the tumors, but might not reflect the exact correspondence (due to possible registration errors), the resulting registrations are regarded as the gold-standard to which I compare in the following. The atlas is registered to: 1) the tumor images (`TUMOR`), 2) the tumor images using cost function masking (`CFM_`$\star$), 3) the quasi-normal images predicted by a network with a quasi-lesion layer (Yang et al., 2016)(`REC_QL`) and 4) the quasi-normal images predicted by the proposed network (`REC_`$\star$). As the gold-standard is obtained through optimization, all the registrations are performed using the same *optimization* model and the *predicted* registrations are not compared. For cost function masking, two experiments are conducted using different masks, one using the groundtruth masks (`CFM_GM`) and one using the predicted masks by the proposed network (`CFM_PM`). Using the predicted masks

(CFM_PM) is to evaluate the performance of cost function masking, when groundtruth (or manually segmented) masks are not available at test time, which is often the case. For the proposed model, networks with (REC_RRS) and without (REC_RR) the segmentation decoder are trained. In addition, for the predicted quasi-normal images, the normal tissue can be kept unchanged by using the predicted segmentation (REC_RRS_PM). The deformation differences are compared between the results obtained by each of the optimization-based registrations and the gold standard registration result.

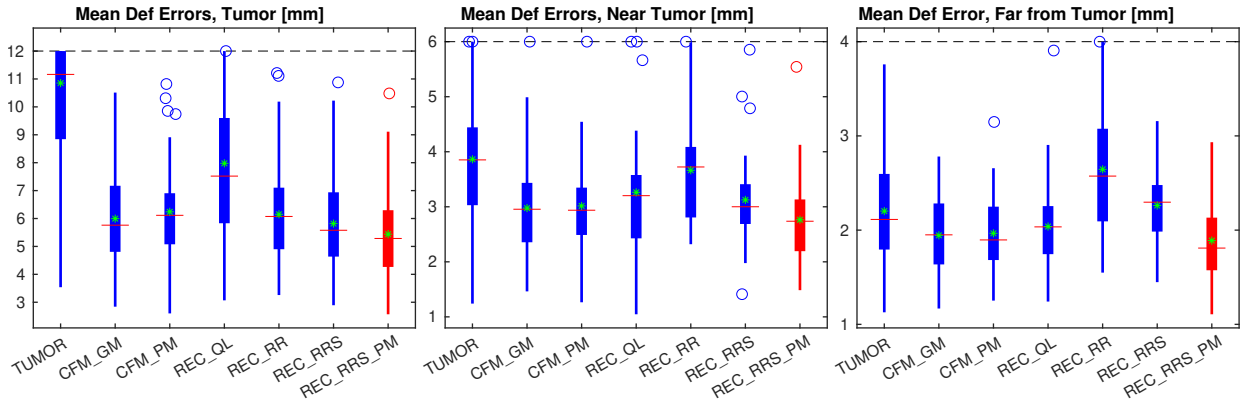


Figure 6.3: Boxplots of mean deformation differences with respect to the gold standard deformations. TUMOR: directly registering to the tumor image; CFM: cost function masking, where _GM and _PM refer to using the groundtruth masks and predicted masks, respectively; REC_⋆s: registering to the reconstructed images, where REC_QL uses the quasi-lesion layer, REC_RR only uses the registration and reconstruction decoders. REC_RRS: proposed network using registration, reconstruction, and segmentation decoders. In addition, REC_RRS_PM (in red) retains the normal region in areas predicted by the masks obtained by the network. Among all methods, the proposed approach (REC_RRS_PM) shows the best performance in all three regions without using groundtruth masks.

Fig. 6.3 shows the results for the pseudo-tumor dataset. For each case, the mean deformation differences are evaluated in three regions: 1) the tumor region, 2) the normal region near the tumor (within 30 mm), and 3) the normal region far from the tumor (over 30 mm). The proposed network performs much better when the segmentation decoder is used (REC_RRS and REC_RR), because of the additional supervision on the segmentation. The network using the quasi-lesion layer (REC_QL) works well in the normal region but performs poorly in the tumor region. This might be because at test time the real tumor region is subject to larger mass effects than what was captured during training, as quasi-lesions can never be introduced inside the actual tumor region.

Compared to cost function masking, the proposed method (REC_RRS_PM), on average, improves by about 0.5 mm in the tumor region when the groundtruth masks are available (CFM_GM) and around 0.8 mm when the groundtruth masks are not available (CFM_PM). In the normal regions, improvements over cost function masking are relatively small, around 0.3 mm. Fig. 6.4 shows three example results of the proposed network for the pseudo-tumor dataset. The 3rd column shows the predicted quasi-normal images, and the 4th column shows the warped image in atlas space.

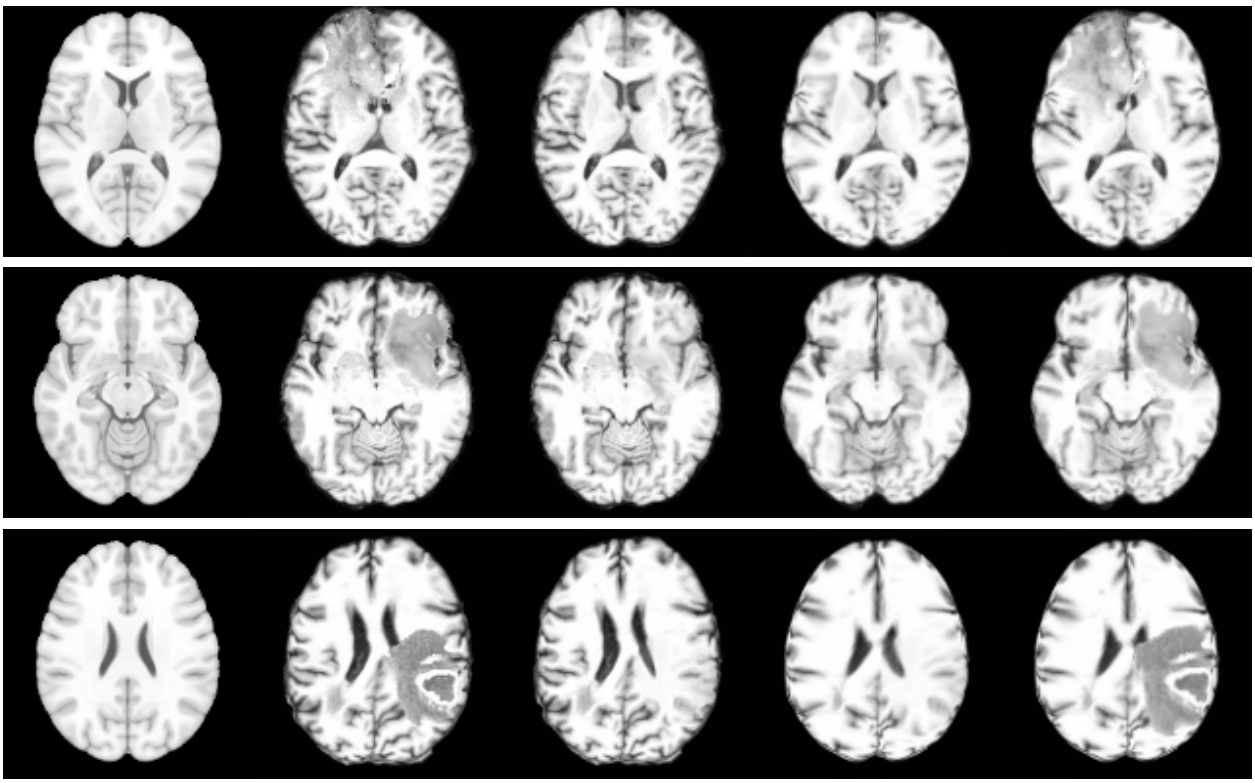


Figure 6.4: Three examples from the pseudo-tumor dataset. The 5 columns show the slices of: (1) the atlas; (2) the tumor image; (3) the reconstructed quasi-normal image, predicted by the proposed network; (4) the warped quasi-normal image by applying the predicted transformation; and (5) the warped tumor image with the same transformation.

6.2.2 3D Real Brain Tumor Dataset

This dataset consists of 22 patients with brain glioblastoma. Each patient has scans from two time-points, one before the surgery (pre-operative) and one after surgery (post-recurrence; after the deformations occurring from tumor resection have relaxed). Each time-point provides

T1w and contrast-enhanced T1w (T1w-CE), T2w and FLAIR MR images. All images are of size $155 \times 240 \times 240$ with isotropic voxels $1 \times 1 \times 1$ mm^3. Only the T1w images are used in the experiments. For each patient, a radiologist placed 10 landmarks near the tumor (within 30 mm) and 10 landmarks far from the tumor (over 30 mm) in both the pre- and post-scans. The proposed network is trained using a subset of the BraTS2019 training data: 120 images for training and 20 images for validation; testing is performed via the glioblastoma dataset. One limitation of using the BraTS data for training is that it only contains pre-operative scans, while the testing images have several post-recurrence scans. To limit dataset variability, a subset of the BraTS training data was selected, which was acquired by one institution and which is similar in acquisition to the test data. Ideally, the test dataset is used for longitudinal registration, i.e., registering between the pre- and post-scans from the same patient. As the network predictions are with respect to an atlas, the following two experiments are conducted:

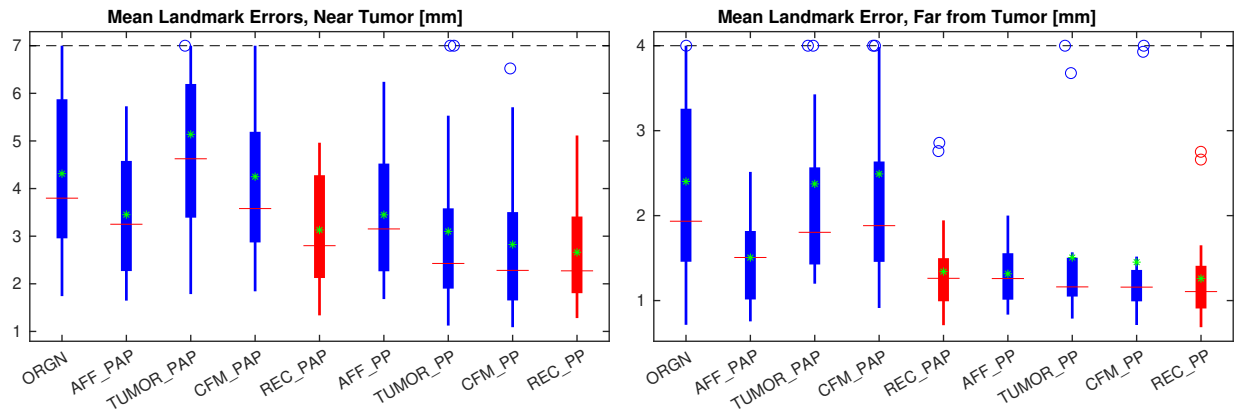


Figure 6.5: Boxplots of mean landmark errors for registration of glioblastoma patients. (`ORGN`) is the landmark differences before registration. The next four are results via the atlas, i.e, pre-atlas-post (`_PAP`); the last four are longitudinal results, i.e., pre-post (`_PP`). Results by the proposed network are compared to affine registration, registration of tumor images, and cost function masking. For atlas registration, the transformation maps predicted by the proposed network are composed. For longitudinal registration, optimization-based vSVF is performed on the predicted quasi-normal images. Results by the proposed network are shown in red.

- **Atlas Registration.** For each patient, both scans are fed into the proposed network and respective transformations to the atlas are obtained. Then the forward map of the pre-scan and the inverse map of the post-scan are composed, resulting in a pre-atlas-post (`REC_PAP`) map. To compare, the vSVF optimization-based atlas-registration are also performed directly

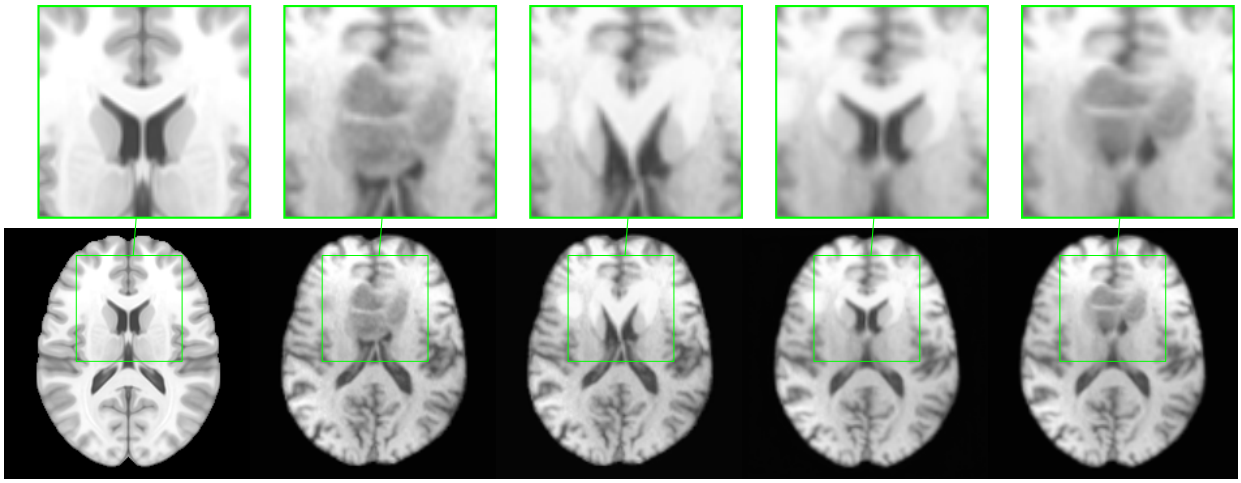


Figure 6.6: One example network result for a brain tumor image. The five columns show: (1) the atlas; (2) the tumor image; (3) the reconstructed quasi-normal image, predicted by the network; (4) the warped quasi-normal image by applying the predicted transformation; and (5) the warped tumor image with the same transformation.

using the tumor images (`TUMOR_PAP`) and with cost function masking (`CFM_PAP`). In both cases,the composited transformation is obtained. Using the resulting transformations, the landmarks are warped from the post-scan to the pre-scan space and the landmark differences are evaluated. As the manual tumor segmentations are not available, predicted masks are used for cost function masking.

- **Longitudinal Registration.** Optimization-based vSVF registrations are performed between reconstructed quasi-normal images of both the scans, predicted by the proposed network (`REC_PP`). Longitudinal registrations directly using tumor images (`TUMOR_PP`) and using cost function masking (`CFM_PP`) are also compared.

Fig. 6.5 shows resulting landmark errors in two different regions for the different registration approaches, including atlas registration and longitudinal registration. When registrations are composed through the atlas, registration errors are much larger than when using the reconstructed quasi-normal images for longitudinal registration. However, the proposed method still shows improvements over cost function masking in both cases. One interesting observation is that the landmark errors through direct tumor-to-atlas registration and cost function masking are even worse

than affine registration when going through the atlas. Finally, Fig. 6.6 shows an example network result for a brain tumor image. The 3rd column shows the predicted quasi-normal images, and the 4th column shows the warped image in atlas space. In the reconstructed image, some contrast differences are observed between the tumor and the normal region. However, as the primary goal is for registration, this difference is not an issue as long as the correspondences can be established between the reconstructed image and the atlas image.

6.2.3 Statistical Analysis and Experiment Settings

Statistical Analysis. To test if the observed differences between the registration approaches are statistically significant, several one-tailed paired Wilcoxon signed-rank tests are performed and the Benjamini-Hochberg (Benjamini and Hochberg, 1995) approach is used to control the false discovery rate at 1%.

- **Pseudo-tumor dataset.** The comparisons are between the proposed method (`REC_RRS_PM`) and cost function masking when groundtruth masks are available or are not available, i.e. `CFM_GM` and `CFM_PM`. The quasi-lesion network (`REC_QL`) and the simplified version of the proposed network when the segmentation decoder is not used (`REC_RR`) are also compared.
- **Real tumor dataset.** Results from proposed network are compared with cost function masking (`REC_PAP vs CFM_PAP` and `REC_PP vs CFM_PP`).

		Tumor	Near	Far
Pseudo	REC_RRS_PM v CFM_GM	1.17e-3(0.80)	5.03e-3(0.84)	0.113(0.63)
	REC_RRS_PM v CFM_PM	5.07e-4(0.70)	5.03e-3(0.80)	0.107(0.68)
	REC_RRS_PM v REC_RR	7.65e-8(0.92)	2.33e-8(0.79)	2.00e-8(0.56)
	REC_RRS_PM v REC_QL	2.21e-6(0.44)	9.51e-4(0.59)	0.010(0.75)
Real	REC_PAP v CFM_PAP	-	3.19e-3(0.55)	6.38e-5(0.90)
	REC_PP vs CFM_PP	-	2.21e-2(0.91)	3.98e-2(0.92)

Table 6.1: p-values and effect-sizes for one-tailed paired Wilcoxon signed-rank tests for multiple comparisons. Effect-sizes are in parentheses. Red text indicates statistically significant results after controlling the false discovery rate at 1%.

Tab. 6.1 shows p-values and effect-sizes for the comparisons. Most of the comparisons are statistically significant, especially in the tumor and near tumor region. Also most of the effect sizes are medium or large (> 0.3).

Experimental Settings. For the vSVF model, `dopri5` (Chen et al., 2018) is used for numerical integration and `L-BFGS` (Liu and Nocedal, 1989) as the optimizer. The momentum field is smoothed with a multi-Gaussian kernel with standard deviations $\{0.05, 0.1, 0.15, 0.2, 0.25\}$ and weights $\{0.067, 0.133, 0.2, 0.267, 0.333\}$. Image space is scaled to $[0, 1]^3$ to allow for easy interpretation of the standard deviations. Three image scales $\{0.25, 0.5, 1.0\}$ are used for optimization, each with 100 iterations. For the proposed network, `AdamW` (Kingma and Ba, 2014; Loshchilov and Hutter, 2017) is used as the optimizer with a cosine annealing learning rate scheduler (Loshchilov and Hutter, 2016). The initial learning rate is 5e-4. The network is trained with a batch size of 4 for 1,000 epochs for the pseudo tumor dataset, and with a batch size of 1 for 500 epochs for the real dataset.

6.3 Conclusion

In this chapter, I proposed a joint registration and reconstruction network. Given a brain image with pathologies, the proposed network simultaneously learns a registration to a common atlas space and a reconstruction of quasi-normal appearance in the atlas space. Experiments show that, as the network disentangles the spatial variation (for example, caused by mass effects or normal anatomical variability) from the appearance differences of the pathology, the reconstructed quasi-normal appearance provides better guidance to the registration. This, in turn, improves registration accuracy.

To the best of my knowledge, this is the first deep learning model that is designed for registration of images with pathologies, which includes joint registration and reconstruction of a quasi-normal image. Compared to the proposed decomposition model, it avoids interleaving of the registration and decomposition: it simultaneously outputs a transformation to the atlas and a reconstruction. While pathology masks are not required at test time, these masks are required during training to learn the reconstruction of normal regions and pathological regions. A more attractive model would

be to automatically learn to reconstruct the normal regions and pathological regions, respectively. For example, LRS or the proposed PCA decomposition may be integrated into the reconstruction loss, thereby eliminating the need for pathological masks. Also, these decompositions may provide better statistical modeling of the normal tissue than the atlas appearance. Hence the reconstruction may be more faithful.

One restriction for the deep learning model is that it is learned on the training dataset. When applied to a new dataset that has a large appearance difference, which is often the case for medical image data, the model may perform poorly. One approach to solve this issue would be dataset-specific fine-tuning. Another possible way is to reduce the appearance difference between the new image and the training dataset, e.g., by applying histogram matching as in Chapter 5.

CHAPTER 7: Discussion

7.1 Summary of Contributions

To summarize, let me first revisit my thesis statement here:

Thesis: *Advanced mathematical models can efficiently and effectively estimate normal image appearance from images with pathologies, thereby helping to improve registration accuracy. Furthermore, a deep regression model allows for fast and accurate registrations of such images.*

The following contributions have been made in this thesis to support my statement:

1. *I proposed an image decomposition framework that reconstructs a quasi-normal image from an image with pathology.*

   Given an image with pathologies, the model (PCA-TV) presented in Chapter 3 decomposes an image with pathologies to a quasi-normal image and its abnormal part. The quasi-normal image is then used for registration. The normal tissue is modeled via PCA of the normal population. The total variation term used in the model captures large pathologies while avoiding image blurring. Similar to the LRS framework (Liu et al., 2014), decomposition steps are alternated with image registration steps to an atlas. Registration experiments have shown that the proposed framework outperforms LRS, and it reaches a similar performance to cost function masking while not requiring prior knowledge of pathology location.

   As the proposed decomposition model pre-registers all normal images to atlas space and only stores one copy of the pre-computed PCA basis, it substantially reduces the memory consumption and speeds up the computation via a GPU implementation.

2. *I developed a patient-specific strategy that models the healthy tissue of a patient in the patient space.*

The proposed patient-specific strategy in Chapter 4 is an extension of the PCA-TV model. It addresses the clinically relevant problem of registration between pre-operative and post-recurrence images for patients with glioblastoma.

Specifically, the post-scan, which is usually free from mass effect, is first used to provide a quasi-normal estimation. Once all normal images are registered to the patient space, a patient-specific PCA is built, which allows for more accurate modeling of the quasi-normal reconstruction for each patient, thereby resulting in improved longitudinal registration accuracy.

In addition to being patient-specific, the proposed model is automatic. It does not require human intervention, neither in the form of tumor seeds or segmentation nor via a complex tumor growth model. It achieves significant accuracy improvements over six conventionally used methods.

3. *I proposed a brain extraction framework that is designed for images with pathologies by jointly modeling the healthy brain issue, pathologies, and non-brain volume.*

   The proposed brain extraction framework in Chapter 5 integrates a spatially distributed sparsity term to capture the skull and surrounding tissue, in addition to the quasi-normal image reconstruction via PCA and pathology identification.

   While the proposed framework is designed for brain image that contains pathologies, it shows improved brain extraction results in datasets with and without pathologies with a fixed set of parameters (without additional tuning or dataset-specific brain templates). On four datasets, the proposed approach either performs best or is among the best methods.

4. *I developed a deep learning model that jointly learns the registration of an image with pathologies to an atlas and the appearance mapping from the pathology to a quasi-normal image.*

   The proposed network in Chapter 6 is more computationally efficient than the decomposition frameworks as it avoids the interleaving of the registration and reconstructions. The network

disentangles the spatial variation from appearance learning, which allows learning the tumor-to-quasi-normal image appearance in atlas space. The reconstructed quasi-normal image is used to drive the registration, which provides more reliable guidance.

Experiments have shown that the proposed network successfully learns to reconstruct quasi-normal image appearance and to register the image to atlas space, and the quasi-normal reconstruction provides better registration results than cost function masking.

7.2 Discussion and Future Work

Finally, in this section, I will present an outlook on possible future work. Some of them may have already been mentioned in the discussion section of each chapter.

7.2.1 Registration of Images with Pathologies

One interesting direction is to use a GAN to learn how to reconstruct a more realistic quasi-normal image, without specifically modeling the normal tissue and pathologies. The generator should create a quasi-normal image from an image with pathologies. Its goal would be to fool the discriminator which learns to distinguish normal images from quasi-normal images. However, in addition to creating a quasi-normal image, the generator is also required to keep the normal region as the original image, which may require additional modifications.

The current runtime bottleneck for the decomposition models is the registration steps. Registration could be substantially improved by replacing with a deep network approach, e.g. the model presented in Chapter 6. Further down the road, the combination of the GAN reconstruction and deep network registration could reduce the runtime into several seconds, once they are trained.

Moreover, regarding the PCA decomposition, it would be natural to use a reconstruction that makes use of a form of Mahalanobis distance (Mahalanobis, 1936). This would then emphasize the eigendirections that explain most of the variance in the training data. Note, however, that the proposed model is relatively insensitive to the number of chosen PCA modes. While different numbers of chosen PCA modes may affect how well the quasi-normal image is reconstructed, the number of PCA modes has only slight effects on the registration results.

A form of robust PCA decomposition is effectively constructed, which prefers outliers that jointly form regions with a low total variation. Instead of modeling the decomposition in this way, it could be interesting to explore an LRS model which uses a partially-precomputed L matrix and gets adapted for a given single image. Such a strategy may allow more efficient computations of the LRS decomposition but would require keeping the entire training dataset in memory (instead of only a basis of reduced dimension). Such an approach could likely also be extended to a form of low-rank-total variation decomposition if desired.

7.2.2 Deep Learning Model

Currently, the network presented in Chapter 6 requires the registration to an atlas space. When applied to longitudinal registration problems, the reconstructed quasi-normal images can be used for direct registration. However, the prediction of the transformation can only work by compositing two atlas transformation for a longitudinal problem. Therefore, extending the work to direct longitudinal registration would also be an interesting direction. The longitudinal changes between the normal tissue may be more subtle, compared to the changes in the pathological region. If the spatial variation, such as mass effect, can be disentangled, the normal tissue appearance should be matched. Regions that have large discrepancies between images would be pathological and, therefore, correspondence needs to be established by reconstruction. On the other hand, spatial variation may not be easily separated without obtaining correspondence between images, especially in the pathological region. Hence, in principle, a joint network for direct longitudinal registration of images with pathologies is possible.

BIBLIOGRAPHY

Aljabar, P., Heckemann, R. A., Hammers, A., Hajnal, J. V., and Rueckert, D. (2009). Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy. *Neuroimage*, 46(3):726–738.

Ashburner, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage*, 38(1):95–113.

Avants, B. B., Epstein, C. L., Grossman, M., and Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26–41.

Avants, B. B., Tustison, N., and Song, G. (2009). Advanced normalization tools (ANTS). *Insight j*, 2(365):1–35.

Bakas, S., Akbari, H., Sotiras, A., Bilello, M., Rozycki, M., Kirby, J. S., Freymann, J. B., Farahani, K., and Davatzikos, C. (2017). Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific data*, 4:170117.

Bakas, S., Reyes, M., Jakab, A., Bauer, S., Rempfler, M., Crimi, A., Shinohara, R. T., Berger, C., Ha, S. M., Rozycki, M., et al. (2018). Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS challenge. *arXiv preprint arXiv:1811.02629*.

Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., and Dalca, A. V. (2018). An unsupervised learning model for deformable medical image registration. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 9252–9260.

Balakrishnan, G., Zhao, A., Sabuncu, M. R., Guttag, J., and Dalca, A. V. (2019). VoxelMorph: a learning framework for deformable medical image registration. *IEEE transactions on medical imaging*, 38(8):1788–1800.

Beg, M. F., Miller, M. I., Trouvé, A., and Younes, L. (2005). Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *International journal of computer vision*, 61(2):139–157.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, 57(1):289–300.

Brant-Zawadzki, M., Gillan, G. D., and Nitz, W. R. (1992). MP RAGE: a three-dimensional, T1-weighted, gradient-echo sequence–initial experience in the brain. *Radiology*, 182(3):769–775.

Brett, M., Leff, A. P., Rorden, C., and Ashburner, J. (2001). Spatial normalization of brain images with focal lesions using cost function masking. *Neuroimage*, 14(2):486–500.

Brock, K. K., Mutic, S., McNutt, T. R., Li, H., and Kessler, M. L. (2017). Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132. *Medical physics*, 44(7):e43–e76.

BROIT, C. (1981). Optimal registration of deformed images. *Doctoral Dissertation, University of Pennsylvania*.

Brown, L. G. (1992). A survey of image registration techniques. *ACM computing surveys (CSUR)*, 24(4):325–376.

Chan, T. F. and Esedoglu, S. (2005). Aspects of total variation regularized L1 function approximation. *SIAM Journal on Applied Mathematics*, 65(5):1817–1837.

Chen, R. T., Rubanova, Y., Bettencourt, J., and Duvenaud, D. K. (2018). Neural ordinary differential equations. In *Advances in neural information processing systems*, pages 6571–6583.

Chitphakdithai, N. and Duncan, J. S. (2010). Non-rigid registration with missing correspondences in preoperative and postresection brain images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 367–374. Springer.

Clausen, C. and Wechsler, H. (2000). Color image compression using PCA and backpropagation learning. *Pattern Recognition*, 33(9):1555–1560.

Cootes, T. F., Edwards, G. J., and Taylor, C. J. (2001). Active appearance models. *IEEE Transactions on pattern analysis and machine intelligence*, 23(6):681–685.

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical research*, 29(3):162–173.

De la Torre, F. and Black, M. J. (2001). Robust principal component analysis for computer vision. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 1, pages 362–369. IEEE.

DO Q, L. (2012). Numerically efficient methods for solving least squares problems.

Doersch, C. (2016). Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*.

Dongarra, J., Gates, M., Haidar, A., Kurzak, J., Luszczek, P., Tomov, S., and Yamazaki, I. (2018). The singular value decomposition: Anatomy of optimizing an algorithm for extreme scale. *SIAM review*, 60(4):808–865.

Doshi, J., Erus, G., Ou, Y., Gaonkar, B., and Davatzikos, C. (2013). Multi-atlas skull-stripping. *Academic radiology*, 20(12):1566–1576.

Drakopoulos, F. and Chrisochoides, N. P. (2016). Accurate and fast deformable medical image registration for brain tumor resection using image-guided neurosurgery. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging & Visualization*, 4(2):112–126.

Draper, B. A., Baek, K., Bartlett, M. S., and Beveridge, J. R. (2003). Recognizing faces with PCA and ICA. *Computer vision and image understanding*, 91(1-2):115–137.

Dupuis, P., Grenander, U., and Miller, M. I. (1998). Variational problems on flows of diffeomorphisms for image matching. *Quarterly of applied mathematics*, pages 587–600.

Eskildsen, S. F., Coupé, P., Fonov, V., Manjón, J. V., Leung, K. K., Guizard, N., Wassef, S. N., Østergaard, L. R., Collins, D. L., Initiative, A. D. N., et al. (2012). BEaST: brain extraction based on nonlocal segmentation technique. *NeuroImage*, 59(3):2362–2373.

Fischer, B. and Modersitzki, J. (1999). Fast inversion of matrices arising in image processing. *Numerical Algorithms*, 22(1):1–11.

Fischer, B. and Modersitzki, J. (2003). Curvature based image registration. *Journal of Mathematical Imaging and Vision*, 18(1):81–85.

Fonov, V. S., Evans, A. C., McKinstry, R. C., Almli, C., and Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, (47):S102.

Givon, L. E., Unterthiner, T., Erichson, N. B., Chiang, D. W., Larson, E., Pfister, L., Dieleman, S., Lee, G. R., van der Walt, S., Menn, B., et al. (2015). scikit-cuda 0.5. 1: a Python interface to GPU-powered libraries.

Goldstein, T., Li, M., Yuan, X., Esser, E., and Baraniuk, R. (2013). Adaptive primal-dual hybrid gradient methods for saddle-point problems. *arXiv preprint arXiv:1305.0546*.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680.

Gooya, A., Biros, G., and Davatzikos, C. (2010). Deformable registration of glioma images using EM algorithm and diffusion reaction modeling. *IEEE transactions on medical imaging*, 30(2):375–390.

Gooya, A., Pohl, K. M., Bilello, M., Biros, G., and Davatzikos, C. (2011). Joint segmentation and deformable registration of brain scans guided by a tumor growth model. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 532–540. Springer.

Gooya, A., Pohl, K. M., Bilello, M., Cirillo, L., Biros, G., Melhem, E. R., and Davatzikos, C. (2012). GLISTR: glioma image segmentation and registration. *IEEE transactions on medical imaging*, 31(10):1941–1954.

Han, X., Bakas, S., Kwitt, R., Aylward, S., Akbari, H., Bilello, M., Davatzikos, C., and Niethammer, M. (2018a). Patient-specific registration of pre-operative and post-recurrence brain tumor MRI scans. In *International MICCAI Brainlesion Workshop*, pages 105–114. Springer.

Han, X., Kwitt, R., Aylward, S., Bakas, S., Menze, B., Asturias, A., Vespa, P., Van Horn, J., and Niethammer, M. (2018b). Brain extraction from normal and pathological images: a joint PCA/image-reconstruction approach. *NeuroImage*, 176:431–445.

Han, X., Shen, Z., Xu, Z., Bakas, S., Akbari, H., Bilello, M., Davatzikos, C., and Niethammer, M. (2020). A deep network for joint registration and reconstruction of images with pathologies. *arXiv preprint arXiv:2008.07628*.

Han, X., Yang, X., Aylward, S., Kwitt, R., and Niethammer, M. (2017). Efficient registration of pathological images: a joint PCA/image-reconstruction approach. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 10–14. IEEE.

Haskins, G., Kruger, U., and Yan, P. (2020). Deep learning in medical image registration: A survey. *Machine Vision and Applications*, 31(1):8.

He, C., Liu, Q., Li, H., and Wang, H. (2010). Multimodal medical image fusion based on IHS and PCA. *Procedia Engineering*, 7:280–285.

Hogea, C., Davatzikos, C., and Biros, G. (2008). An image-driven parameter estimation problem for a reaction–diffusion glioma growth model with mass effects. *Journal of mathematical biology*, 56(6):793–825.

Huizinga, W., Poot, D. H., Guyader, J.-M., Klaassen, R., Coolen, B. F., van Kranenburg, M., Van Geuns, R., Uitterdijk, A., Polfliet, M., Vandemeulebroucke, J., et al. (2016). PCA-based groupwise image registration for quantitative MRI. *Medical image analysis*, 29:65–78.

Iglesias, J. E., Liu, C.-Y., Thompson, P. M., and Tu, Z. (2011). Robust brain extraction across datasets and comparison with publicly available methods. *IEEE Transactions on Medical Imaging*, 30(9):1617–1634.

Irimia, A., Wang, B., Aylward, S. R., Prastawa, M. W., Pace, D. F., Gerig, G., Hovda, D. A., Kikinis, R., Vespa, P. M., and Van Horn, J. D. (2012). Neuroimaging of structural pathology and connectomics in traumatic brain injury: Toward personalized outcome prediction. *NeuroImage: Clinical*, 1(1):1–17.

Jenkinson, M., Beckmann, C. F., Behrens, T. E., Woolrich, M. W., and Smith, S. M. (2012). Fsl. *Neuroimage*, 62(2):782–790.

Jolliffe, I. (2003). Principal component analysis. *Technometrics*, 45(3):276.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kingma, D. P. and Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Kistler, M., Bonaretti, S., Pfahrer, M., Niklaus, R., and Büchler, P. (2013). The virtual skeleton database: an open access repository for biomedical research and collaboration. *Journal of medical Internet research*, 15(11):e245.

Kleesiek, J., Urban, G., Hubert, A., Schwarz, D., Maier-Hein, K., Bendszus, M., and Biller, A. (2016). Deep MRI brain extraction: A 3D convolutional neural network for skull stripping. *NeuroImage*, 129:460–469.

Kwon, D., Niethammer, M., Akbari, H., Bilello, M., Davatzikos, C., and Pohl, K. M. (2014). PORTR: Pre-operative and post-recurrence brain tumor registration. *IEEE transactions on medical imaging*, 33(3):651–667.

Kwon, D., Zeng, K., Bilello, M., and Davatzikos, C. (2015). Estimating patient specific templates for pre-operative and follow-up brain tumor registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 222–229. Springer.

LaMontagne, P. J., Benzinger, T. L., Morris, J. C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J., Moulder, K., Vlassenko, A., et al. (2019). OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. *medRxiv*.

LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551.

Lee, C., Langen, K. M., Lu, W., Haimerl, J., Schnarr, E., Ruchala, K. J., Olivera, G. H., Meeks, S. L., Kupelian, P. A., Shellenberger, T. D., et al. (2008). Assessment of parotid gland dose changes during head and neck cancer radiotherapy using daily megavoltage computed tomography and deformable image registration. *International Journal of Radiation Oncology* Biology* Physics*, 71(5):1563–1571.

Li, S., Fevens, T., Krzyżak, A., and Li, S. (2006). Automatic clinical image segmentation using pathological modeling, PCA and SVM. *Engineering Applications of Artificial Intelligence*, 19(4):403–410.

Lin, Z., Chen, M., and Ma, Y. (2010). The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*.

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., and Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88.

Liu, D. C. and Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization. *Mathematical programming*, 45(1-3):503–528.

Liu, X., Niethammer, M., Kwitt, R., McCormick, M., and Aylward, S. (2014). Low-rank to the rescue–atlas-based analyses in the presence of pathologies. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 97–104. Springer.

Liu, X., Niethammer, M., Kwitt, R., Singh, N., McCormick, M., and Aylward, S. (2015). Low-rank atlas image analyses in the presence of pathologies. *IEEE transactions on medical imaging*, 34(12):2583–2591.

Loshchilov, I. and Hutter, F. (2016). Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*.

Loshchilov, I. and Hutter, F. (2017). Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.

Lowekamp, B. C., Chen, D. T., Ibáñez, L., and Blezek, D. (2013). The design of SimpleITK. *Frontiers in neuroinformatics*, 7:45.

Mahalanobis, P. C. (1936). On the generalized distance in statistics. National Institute of Science of India.

Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., and Buckner, R. L. (2007). Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *Journal of cognitive neuroscience*, 19(9):1498–1507.

Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al. (2014). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE transactions on medical imaging*, 34(10):1993–2024.

Milano, M. T., Katz, A. W., and Okunieff, P. (2010). Patterns of recurrence after curative-intent radiation for oligometastases confined to one organ. *American journal of clinical oncology*, 33(2):157–163.

Modat, M., Cash, D. M., Daga, P., Winston, G. P., Duncan, J. S., and Ourselin, S. (2014). Global image registration using a symmetric block-matching approach. *Journal of Medical Imaging*, 1(2):024003.

Modat, M., Daga, P., Cardoso, M. J., Ourselin, S., Ridgway, G. R., and Ashburner, J. (2012). Parametric non-rigid registration using a stationary velocity field. In *2012 IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*, pages 145–150. IEEE.

Modat, M., Ridgway, G. R., Taylor, Z. A., Lehmann, M., Barnes, J., Hawkes, D. J., Fox, N. C., and Ourselin, S. (2010). Fast free-form deformation using graphics processing units. *Computer methods and programs in biomedicine*, 98(3):278–284.

Modersitzki, J. (2004). *Numerical methods for image registration*. Oxford University Press on Demand.

Muschelli, J., Ullman, N. L., Mould, W. A., Vespa, P., Hanley, D. F., and Crainiceanu, C. M. (2015). Validated automatic brain extraction of head CT images. *Neuroimage*, 114:379–385.

Nickolls, J., Buck, I., Garland, M., and Skadron, K. (2008). Scalable parallel programming with CUDA. *Queue*, 6(2):40–53.

Niethammer, M., Hart, G. L., Pace, D. F., Vespa, P. M., Irimia, A., Van Horn, J. D., and Aylward, S. R. (2011). Geometric metamorphosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 639–646. Springer.

Niethammer, M., Kwitt, R., and Vialard, F.-X. (2019). Metric learning for image registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8463–8472.

Nwankpa, C., Ijomah, W., Gachagan, A., and Marshall, S. (2018). Activation functions: Comparison of trends in practice and research for deep learning. *arXiv preprint arXiv:1811.03378*.

Omuro, A. and DeAngelis, L. M. (2013). Glioblastoma and other malignant gliomas: a clinical review. *Jama*, 310(17):1842–1850.

Osher, S., Burger, M., Goldfarb, D., Xu, J., and Yin, W. (2005). An iterative regularization method for total variation-based image restoration. *Multiscale Modeling & Simulation*, 4(2):460–489.

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1):62–66.

Ou, Y., Sotiras, A., Paragios, N., and Davatzikos, C. (2011). DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting. *Medical image analysis*, 15(4):622–639.

Ourselin, S., Roche, A., Subsol, G., Pennec, X., and Ayache, N. (2001). Reconstructing a 3D structure from serial histological sections. *Image and vision computing*, 19(1-2):25–31.

Pace, D. F., Niethammer, M., and Aylward, S. R. (2011). Sliding geometries in deformable image registration. In *International MICCAI Workshop on Computational and Clinical Challenges in Abdominal Imaging*, pages 141–148. Springer.

Price, S. J., Jena, R., Burnet, N. G., Carpenter, T. A., Pickard, J. D., and Gillard, J. H. (2007). Predicting patterns of glioma recurrence using diffusion tensor imaging. *European radiology*, 17(7):1675–1684.

Provenzale, J. M., Mukundan, S., and Barboriak, D. P. (2006). Diffusion-weighted and perfusion MR imaging for brain tumor characterization and assessment of treatment response. *Radiology*, 239(3):632–649.

Richardson, M. (2009). Principal component analysis. *URL: http://people. maths. ox. ac. uk/richardsonm/SignalProcPCA. pdf (last access: 3.5. 2013). Aleš Hladnik Dr., Ass. Prof., Chair of Information and Graphic Arts Technology, Faculty of Natural Sciences and Engineering, University of Ljubljana, Slovenia ales. hladnik@ ntf. uni-lj. si*, 6:16.

Risser, L., Vialard, F.-X., Wolz, R., Murgasova, M., Holm, D. D., and Rueckert, D. (2011). Simultaneous multi-scale registration using large deformation diffeomorphic metric mapping. *IEEE transactions on medical imaging*, 30(10):1746–1759.

Rudin, L. I., Osher, S., and Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268.

Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L., Leach, M. O., and Hawkes, D. J. (1999). Nonrigid registration using free-form deformations: application to breast MR images. *IEEE transactions on medical imaging*, 18(8):712–721.

Ségonne, F., Dale, A. M., Busa, E., Glessner, M., Salat, D., Hahn, H. K., and Fischl, B. (2004). A hybrid approach to the skull stripping problem in MRI. *Neuroimage*, 22(3):1060–1075.

Shattuck, D. W. and Leahy, R. M. (2002). BrainSuite: an automated cortical surface identification tool. *Medical image analysis*, 6(2):129–142.

Shattuck, D. W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K. L., Poldrack, R. A., Bilder, R. M., and Toga, A. W. (2008). Construction of a 3D probabilistic atlas of human cortical structures. *Neuroimage*, 39(3):1064–1080.

Shattuck, D. W., Sandor-Leahy, S. R., Schaper, K. A., Rottenberg, D. A., and Leahy, R. M. (2001). Magnetic resonance image tissue classification using a partial volume model. *NeuroImage*, 13(5):856–876.

Shen, Z., Han, X., Xu, Z., and Niethammer, M. (2019a). Networks for joint affine and non-parametric image registration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4224–4233.

Shen, Z., Vialard, F.-X., and Niethammer, M. (2019b). Region-specific diffeomorphic metric mapping. In *Advances in Neural Information Processing Systems*, pages 1096–1106.

Shu, Z., Sahasrabudhe, M., Alp Guler, R., Samaras, D., Paragios, N., and Kokkinos, I. (2018). Deforming autoencoders: Unsupervised disentangling of shape and appearance. In *Proceedings of the European conference on computer vision (ECCV)*, pages 650–665.

Singh, N., Hinkle, J., Joshi, S., and Fletcher, P. T. (2013). A vector momenta formulation of diffeomorphisms for improved geodesic regression and atlas construction. In *2013 IEEE 10th International Symposium on Biomedical Imaging*, pages 1219–1222. IEEE.

Sled, J. G., Zijdenbos, A. P., and Evans, A. C. (1998). A nonparametric method for automatic correction of intensity nonuniformity in MRI data. *IEEE transactions on medical imaging*, 17(1):87–97.

Smith, S. M. (2002). Fast robust automated brain extraction. *Human brain mapping*, 17(3):143–155.

Smith, S. M., Zhang, Y., Jenkinson, M., Chen, J., Matthews, P., Federico, A., and De Stefano, N. (2002). Accurate, robust, and automated longitudinal and cross-sectional brain change analysis. *Neuroimage*, 17(1):479–489.

Sonka, M. and Fitzpatrick, J. M. (2000). Handbook of medical imaging(Volume 2, Medical image processing and analysis). SPIE- The international society for optical engineering.

Speier, W., Iglesias, J. E., El-Kara, L., Tu, Z., and Arnold, C. (2011). Robust skull stripping of clinical glioblastoma multiforme data. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 659–666. Springer.

Toga, A. W. and Thompson, P. M. (2001). The role of image registration in brain mapping. *Image and vision computing*, 19(1-2):3–24.

Turk, M. and Pentland, A. (1991). Face recognition using eigenfaces. In *Proceedings. 1991 IEEE computer society conference on computer vision and pattern recognition*, pages 586–587.

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., and Gee, J. C. (2010). N4ITK: improved N3 bias correction. *IEEE transactions on medical imaging*, 29(6):1310–1320.

Vercauteren, T., Pennec, X., Perchant, A., and Ayache, N. (2009). Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage*, 45(1):S61–S72.

Voulodimos, A., Doulamis, N., Doulamis, A., and Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, 2018.

Wilcoxon, F., Katti, S., and Wilcox, R. A. (1970). Critical values and probability levels for the Wilcoxon rank sum test and the Wilcoxon signed rank test. *Selected tables in mathematical statistics*, 1:171–259.

Wilks, D. S. (2011). Principal component (EOF) analysis. In *International Geophysics*, volume 100, pages 519–562. Elsevier.

Wollny, G. and Kruggel, F. (2002). Computational cost of nonrigid registration algorithms based on fluid dynamics [MRI time series application]. *IEEE Transactions on Medical Imaging*, 21(8):946–952.

Worth, A. (1996). The internet brain segmentation repository (IBSR).

Wright, J., Ganesh, A., Rao, S., Peng, Y., and Ma, Y. (2009). Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In *Advances in neural information processing systems*, pages 2080–2088.

Yang, X., Han, X., Park, E., Aylward, S., Kwitt, R., and Niethammer, M. (2016). Registration of pathological images. In *International Workshop on Simulation and Synthesis in Medical Imaging*, pages 97–107. Springer.

Yang, X., Kwitt, R., Styner, M., and Niethammer, M. (2017). Quicksilver: Fast predictive image registration–a deep learning approach. *NeuroImage*, 158:378–396.

Yeghiazaryan, V. and Voiculescu, I. (2015). An overview of current evaluation methods used in medical image segmentation. *Department of Computer Science, University of Oxford*.

Young, T., Hazarika, D., Poria, S., and Cambria, E. (2018). Recent trends in deep learning based natural language processing. *ieee Computational intelligenCe magazine*, 13(3):55–75.

Zitova, B. and Flusser, J. (2003). Image registration methods: a survey. *Image and vision computing*, 21(11):977–1000.